ChromID identifies the protein interactome at chromatin marks

ChromID reveals the proteome composition of key chromatin states in murine stem cells.

Rodrigo Villaseñor (1), Ramon Pfaendler (1), Christina Ambrosi (1,2), Stefan Butz (1,2), Sara Giuliani (1), Elana Bryan (3), Thomas W. Sheahan (3), Annika Gable (4,2), Nina Schmolka (1), Massimiliano Manzo (1,2), Joël Wirz (1), Christian Feller (5), Christian von Mering (4), Ruedi Aebersold (5), Philipp Voigt (3), and Tuncay Baubec (1)

1. Department of Molecular Mechanism of Disease, University of Zurich, Zurich, Switzerland
2. Life Science Zurich Graduate School, University of Zurich and ETH Zurich, Zurich, Switzerland
3. Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK
4. Institute of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich, Zurich, Switzerland
5. Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Switzerland and Faculty of Science, University of Zurich, Switzerland

Correspondence: tuncay.baubec@uzh.ch

Abstract

Chromatin modifications instruct genome function through spatiotemporal recruitment of regulatory factors to the genome. However, how these modifications define the proteome composition at distinct chromatin states remains to be fully characterized. Here, we made use of natural protein domains as modular building blocks to develop engineered chromatin readers (eCRs) selective for histone and DNA modifications. By stably expressing eCRs in mouse embryonic stem cells and measuring their subnuclear localisation, genomic distribution and histone-PTM-binding preference, we first demonstrate their applicability as selective chromatin binders in living cells. Finally, we exploit the binding specificity of eCRs to establish ChromID, a new method for chromatin-dependent proteome identification based on proximity biotinylation. We use ChromID to reveal the proteome at distinct chromatin modifications in mouse stem cells, and by using a synthetic dual-modification reader, we furthermore uncover the protein composition at bivalent promoters marked by H3K4me3 and H3K27me3. These results highlight the applicability of ChromID as novel method to obtain a detailed view of the protein interaction network determined by the chemical language on chromatin.

Keywords:

Chromatin, Epigenetics, Synthetic Biology, Systems Biology, Chromatin Proteomics

Introduction

Chromatin and numerous chemical modifications on histones and DNA play critical roles in organismal development and human health [1]. These modifications are recognised by specialised reader domains that exist in numerous regulatory proteins and multiprotein complexes [2,3]. Depending on the presence and composition of modifications at genomic sites, regulatory factors can associate with chromatin in a spatiotemporal manner [4]. However, a major challenge in the field remains to understand how this chemical language on chromatin defines the protein interactome of the genome.

In recent years, proteomics-based assays helped to measure the chromatin interaction preferences of numerous proteins and to discover their affinity to chromatin marks. Current methods probe the cellular proteome using synthetic histone peptides, methylated DNA probes or *in-vitro*-reconstituted nucleosomes [5-9]. In addition, proteomics of specific genomic segments can be achieved using enrichment via antibodies, DNA sequence-specific probes or more recently, via engineered dCas9-fusion proteins [10-14]. While these studies have greatly enhanced our current knowledge about interactions between proteins and chromatin marks, the available methods rely on artificial chromatin, protein-protein crosslinking, or methods that require access to the underlying DNA, leading to chromatin disruption. Therefore, novel approaches are required that enable detection of dynamic interactions between proteins and physiological chromatin in living cells.

Here, we developed chromatin-dependent proteome identification (ChromID) to identify the local proteome composition based on the physiological readout of individual and combinatorial chromatin marks. Towards this, we used the reader domains from well-established chromatin regulators as modules to build engineered chromatin readers (eCRs). We first quantified and functionally validated the genome-wide binding and histone-PTM interaction preferences of individual eCRs towards DNA methylation, H3K9me3, H3K4me3

and H3K27me3, demonstrating their applicability as selective binders in mouse stem cells. Finally, we utilised the specificity of eCRs to recruit promiscuous biotin ligases to detect proteins associated with these individual chromatin modifications in mouse embryonic stem cells, revealing similarities and differences in the proteome composition between these marks. By coupling ChromID to a synthetic dual-modification reader, we furthermore detect proteins associated with genomic regions marked by the bivalent modification H3K4me3 and H3K27me3.

Results

Generation and characterisation of engineered chromatin readers in mouse embryonic stem cells

We first assembled well-characterised chromatin reader domains into synthetic reporter proteins to test their affinity and specificity for individual chromatin modifications in living cells. Towards this, we used the chromo domains specific for H3K27me3 from CBX7 and Drosophila Polycomb (dPC) [15,16], the H3K9me3-specific chromodomain from CBX1 [17,18], the Phd domain specific for H3K4me3 from TAF3 [19], and the MBD domains from the DNA methylation readers MBD1 and MeCP2 [20,21] (Fig. 1a). cDNA sequences were assembled either as single- or dual-domain constructs into a protein expression cassette containing a biotin acceptor site for biochemical purification, a nuclear localisation signal (NLS), and eGFP for live imaging and detection (Fig. 1a and Supplementary Fig. 1a). All constructs were integrated to a defined site in the mouse genome via Recombinase-Mediated Cassette Exchange RMCE [22], enabling fast generation of stable mESC lines expressing the proteins from the same genomic location and under the control of the same promoter (Fig. 1b). Measurements of eGFP fluorescence and protein levels indicated that all generated cell lines display stable and homogenous expression of the introduced engineered Chromatin Readers (eCR) at intermediate protein levels (Supplementary Fig. 1b-d). We next performed *in vitro* differentiation of mESC to glutamatergic neurons to test if the presence of eCRs interferes with biological processes relevant for cellular identity and function. All cell lines successfully differentiated to mature neurons, and we could not observe any differentiation defects in presence or absence of the eCRs, suggesting that their stable expression does not interfere with cellular processes (Supplementary Fig. 1e-g).

Live imaging of stable cell lines with single chromodomain eCRs targeting histone methylation showed a diffuse nuclear localisation with accumulation in nucleoli, similar to the

eGFP control lacking reader domains. In contrast, eCRs containing two chromodomains showed a defined pattern with signals for the CBX1-2xChromo eCR at DAPI-dense chromocenters and the CBX7- and dPC2-2xChromo eCRs forming discrete subnuclear aggregates at the nuclear periphery and around nucleoli (Fig. 1c and Supplementary Fig. 2a). These localisation patterns are identical to what has been reported for the subnuclear distribution of H3K9me3 and H3K27me3, respectively [23,24] (Supplementary Fig. 2b). In contrast to the chromodomains, the single and dual TAF3 Phd-domain eCR showed a homogenous signal throughout the entire nucleus (Supplementary Fig. 2a) as previously reported for H3K4me3 distribution using antibodies [24]. eCRs containing single MBD domains from MBD1 or MeCP2 co-localised with DNA-methylated, DAPI-dense chromocenters, similar to their corresponding full-length proteins (Supplementary Fig. 2a) [22]. Furthermore, live-cell imaging of the eGFP-tagged eCRs enabled us to explore their localisation during cell cycle progression and along the condensed M-phase chromosomes (Supplementary Fig. 2c-e and Supplementary Videos).

The results obtained from single and double domain eCRs specific for histone modifications indicate that one domain is not sufficient to promote localisation and that multivalent interactions are required. To further validate this, we made use of reconstituted nucleosomes carrying H3K4me3 or H3K27me3 marks on both histone tails. Pulldown experiments using recombinant single- or double- domain eCRs indicate robust interactions only for dual domains, but not for single-domain eCRs, supporting the necessity of multivalent interactions for stable binding of eCRs to histone-PTMs *in vivo* (Supplementary Fig. 2f).

<u>Functional analysis validates the interaction preference of eCRs with specific chromatin modifications.</u>

Next, we explored the genome-wide binding patterns of all eCRs by biotin-ChIP-seq [22]. By visual inspection of the binding tracks we observed eCR-specific signals corresponding to the distribution of target histone modifications and DNA methylation, indicating correct localisation to these marks (Fig. 1d-f and Supplementary Fig. 3a-b). Their selective binding preference to chromatin modifications was also confirmed by genome-wide enrichments and direct comparison to histone modifications, DNA methylation and endogenous reader proteins (Fig. 1g, Supplementary Fig. 3c-d and 4a-d). The signals obtained from individual eCR datasets indicated their clear distinction in binding to genomic elements modified by the corresponding target modifications (Supplementary Fig. 5a-d), where eCRs specific for H3K4me3 preferentially associated with gene promoters, H3K9me3 readers with repetitive elements, and DNA methylation readers with methyl-CpG-dense exons (Supplementary Fig. 5e). Notably, and in accordance with live-cell imaging, only experiments using eCRs with two histone-PTM reader domains resulted in detectable binding signals (Supplemental Fig. 3a-b and 5f), highlighting again the necessity for multivalent interactions for stable target engagement.

To investigate the specificity of eCRs we introduced mutations to the reader domains known to disrupt binding: CBX1-W42A [25], CBX7-W35A [26], MBD1-R22A [27] and furthermore, various Rett syndrome mutations in the MeCP2 MBD (R106W, R133C, T158M [28]) (Supplementary Fig. 6a). In all tested instances, we observed that mutations led to a partial or complete disruption of subnuclear localisation (Supplementary Fig. 6b), as well as loss of genome-wide binding to chromatin modifications (Fig. 2a-b, d and Supplementary Fig. 6c). The same disruption of localisation was observed for wild type readers in absence of the respective chromatin marks, highlighting that binding is fully dependent on the target modification. This dependency was shown by loss of binding of MeCP2-1xMBD in cell lines lacking DNA methylation (*Dnmt*-TKO), and CBX7-2xChromo in cell lines lacking H3K27me3 (*Eed*-KO) (Fig. 2c and Supplementary Fig. 6d-f).

Next, we employed mass spectrometry as an orthogonal approach to identify the histone PTMs that are preferentially bound by the eCRs in living cells. Towards this, we detected and quantified the modifications on histones enriched in ChIP experiments using a synthetic reference peptide library including 87 individual and combined marks on histone tails from H2A, H3 and H4 [29] (Supplemental Fig. 7a and b). Overall, the enriched histone PTMs reflect the genome-wide correlations described above, further corroborating the specific affinity of the reader domains for their substrates in living cells (Fig. 2e). In case of the TAF3-2xPhd eCR, we detect histone H3 tails that carry di- and tri- but not mono-methylation marks at the lysine 4 residue and furthermore acetylated H3 and H4 (K9, K14 on histone H3 and K5, K8, K12, K16 on histone H4) (Fig. 2e-f and Supplementary Fig. 7c). Histone H3 tails containing methylated K9, K27 or K36 residues were generally depleted in the TAF3-2xPhd eCR pulldowns (Fig. 2e). In contrast, CBX1-2xChromo eCR-enriched histone tails predominantly carry the H3K9 tri-methyl mark as well as H4K20me3 (Figure 2e-f), a modification co-existing with H3K9me3 at repetitive heterochromatin [30] (Supplementary Fig. 7d-e). In addition, we also detect that H3S10-phosphorylation prevents binding of the CBX1-eCR to H3K9me3, as previously reported for CBX1 *in vitro* [31] (Fig. 2e,g).

Taken together, these experiments validate the target specificity of the introduced eCRs to chromatin marks. Furthermore, the obtained results highlight the suitability of eCRs as multi-purpose cellular probes to detect the distribution of chromatin modifications in living cells by live imaging, genomics and proteomics.

ChromID reveals the proteome composition at H3K9- and DNA-methylated sites via eCR-mediated proximity biotin labelling

Having fully characterised the *in vivo* binding specificity of eCRs, we wanted to exploit their genomic localisation to detect the proteome composition at distinct chromatin

189 modifications via proximity biotin ligation (Fig. 3a). To this end, we tested three different
190 promiscuous biotin ligases for their labelling efficiency during 24 hours in murine ES cells:
191 BirA* R118G-mutant [32], BioID2 [33] and BASU [34] (Supplementary Fig. 8a-b). BioID2 and BASU
192 showed the highest labelling efficiency under these conditions, therefore we used these
193 ligases in combinations with the specific H3K9me3-reader to establish the optimal conditions
194 using quantitative label-free LC-MS/MS. The protein-interactome of this mark has been well-
195 described in mammalian cells [5,6], and served as a proof of concept to define optimal settings
196 for ChromID (Supplementary Fig. 8c-e and Methods). Finally, based on the signal to noise
197 ratio we chose BASU with 12 hours biotin-treatment followed by highly stringent washing with
198 SDS as the most-optimal condition.
199
200 Using these conditions, we identified 58 high-confidence H3K9me3-associated
201 proteins enriched for Gene Ontology terms linked to pericentric or telomeric heterochromatin,
202 confirming proteins found in other proteomic approaches and forward screens [5,6,35] [36](Fig. 3b-
203 c, Supplementary Fig. 8e and Table 1). Found factors include the H3K9 methyltransferases
204 SETDB1, EHMT1 and EHMT2 [37,38], the HUSH complex component MPP8 [39], the chromatin
205 remodeller ATRX [40], MeCP2 and UHRF1 [41], [42]. Besides these factors, our method enabled us
206 to identify zinc finger proteins, which have been linked to heterochromatin (POGZ, WIZ [43,44]),
207 and multiple instances that have not been characterised in the context of heterochromatin
208 (CASZ1, ZNF24, ZNF292, ZNF512B, ZNF518B, ZNF280B and ZNF280D). To test if the
209 newly-identified proteins localize to H3K9me3-marked chromatin, we further validated the
210 localization of the endogenous ZNF280D protein. For this, we endogenously tagged ZNF280D
211 in mouse ES cells and performed biotin ChIP-seq (Supplementary Fig. 9a-b). Genome-wide
212 binding of ZNF280D shows a strong localisation preference to H3K9me3 sites, confirming that
213 ChromID indeed reveals proteins associated with specific chromatin marks (Supplementary
214 Fig. 9c-d).
215

We next used ChromID in conjunction with the 5-methyl-CpG-specific eCR (MBD1-1xMBD) resulting in the identification of proteins associated with DNA methylation such as DNMT1 and UHRF1 [45] and proteins enriched for Gene Ontology terms related to heterochromatin or recognition of DNA replication (Fig. 3d-e, Supplementary Fig. 10a and Table 2). Besides known factors, we also observed several novel factors that have not been associated with DNA methylation in ES cells, such as TIF1A (also known as TRIM24), CASZ1, ZNF512B or TEAD1 (Fig. 3d). The latter was recently found to bind to methylated DNA in HT-SELEX experiments [46]. To test the specificity of these interactions for DNA methylation readout, we repeated these experiments using the mutant 5mC-reader (MBD1-1xMBD-R22A) fused to BASU. We did not detect any significantly enriched proteins with the mutant 5mC-reader, suggesting that the identified proteins associate with DNA methylation (Fig. 3d and Supplementary Fig. 10a-c).

Engineered readout of combinatorial histone PTMs enables identification of proteins associated with monovalent and bivalent chromatin.

Nucleosomes bivalently modified by H3K4me3 and H3K27me3 are found at developmental gene promoters and are thought to poise their expression for timely activation [47-49]. Addressing the genomic distribution and/or protein composition of bivalently modified sites and other combinatorial modifications has been a major challenge due to lack of tools that enable simultaneous detection of both marks. To overcome this limitation, we first characterised synthetic readers engineered for simultaneous detection of H3K4me3 and H3K27me3 on the same nucleosome. eCRs containing the CBX7-Chromodomain or the dPC-Chromodomain fused to the TAF3-Phd domain were stably expressed in ES cells as described above (Supplementary Fig. 11a-b). Genome-wide binding analysis indicates preferential binding of these bivalent eCRs to genomic sites marked by both H3K4me3 and H3K27me3 modifications, while regions containing either H3K4me3 or H3K27me3 were not enriched to the same levels (Fig. 4a-c and Supplementary Fig. 11c-e). Monovalent eCRs with affinity to

H3K4me3 or H3K27me3 only, showed reduced enrichments to bivalent regions, while being predominantly recruited to sites modified by H3K4me3 or H3K27me3, respectively (Fig. 4a-b and Supplementary Fig. 11c,f). To test the requirement of both domains for the observed binding, we introduced mutations in either the TAF3-Phd (DW890/891AA [19]) or the CBX7-Chromo domains (W35A [26]) of the bivalent reader. We observe loss of binding at bivalent sites for both mutant variants (Supplemental Fig. 12a-c). To further evaluate the requirement of both histone modifications for recruitment of the bivalent readers, we have introduced the TAF3-CBX7-bivalent eCR to *Eed*-KO ES cells lacking H3K27me3. In absence of H3K27me3, the bivalent reader fails to bind to the genome (Supplemental Fig. 12d-e), further supporting the finding that its binding is dependent on multivalent readout of both modifications by the two reader domains. Importantly, mutations in the reader domains or absence of H3K27me3 lead to complete loss of binding to the genome with no additional binding observed at H3K4me3- or H3K27me3-marked sites, further indicating that the bivalent readers are specific for bivalently modified nucleosomes. Taken together, the modular architecture of eCRs opens new possibilities to study and manipulate combinatorial modifications in living cells.

The differences in genomic binding observed for the monovalent and bivalent eCRs encouraged us to perform ChromID with eCRs specific to H3K4me3, H3K27me3 and bivalent nucleosomes. In total, 136 unique proteins that directly or indirectly interact with the chromatin marks were found significantly enriched across these three datasets (Supplementary Fig. 13a-b and Table 1). A total of 125 proteins were detected at H3K4me3 (TAF3-eCR), enriching for GO terms related to transcriptional regulation and H3K4me3 (Fig. 5a), including several transcription factors, bromodomain proteins, histone modifier and chromatin remodelling complexes, as well as members of the Transcription Factor IID (TFIID), Integrator-, Mediator- and Super Elongation-complexes (Supplementary Fig. 13a-b and Table 1). Notably, we also detect proteins involved in co-transcriptional processes such as the RNA-specific adenosine deaminase ADAR1 and the histone mRNA 3' end processing factor CASP8AP2/FLASH [50] to

be associated with H3K4me3, which we confirmed by comparing genomic co-localisation of FLASH and H3K4me3 at transcribed histone genes (Fig. 5b).

The H3K27me3-reader enabled us to identify 20 high-confidence hits, enriching for GO terms associated with Polycomb repressive complexes and histone methyltransferases (Fig. 5a, Supplementary Fig. 13a-b and Table 1). Among those hits we observed well-studied subunits of PRC1 and PRC2 (RING2, EZH2, MTF2 and JARD2). Notably, we also identified factors involved in H3K9 methylation like SETDB1 or the zinc finger proteins WIZ and ZNF518B, suggesting a potential crosstalk between proteins bound at H3K9me3 and H3K27me3 sites. Importantly, this is not due to unspecific localisation of the H3K27me3 readers to H3K9me3 or vice versa, since we do not observe this cross-reactivity from our ChIP-seq data (Supplementary Fig. 13c). Furthermore, by performing ChromID with the H3K27me3-specific readers in ES cells lacking H3K27me3 (*Eed*-KO), we would expect that such unspecific interactions would persist. However, we fail to detect any enriched proteins, indicating that the reported interactions indeed originate from H3K27me3 sites (Supplementary Fig. 13d).

Finally, the combinatorial recognition of bivalent H3K4me3 and H3K27me3 loci by the CBX7-TAF3-eCR enabled us to discover 33 high-confidence factors associated with bivalent chromatin, enriched in GO terms related to transcriptional activation and repression (Fig. 5a, Supplementary Fig. 13a-b and Table 1). These included catalytic subunits or components of the MLL1/MLL2, the NSL histone acetyltransferase and the TFIID basal transcription factor complex, although TFIID components were detected at lower levels compared to results obtained with the monovalent H3K4me3 reader. Other factors include enhancer of Polycomb homolog (EPC1 and EPC2) and components of the NuA4 histone acetyltransferase (HAT) complex [51,52]. Corroborating our findings, a recent study mapped the catalytic subunit of the NuA4 complex (TIP60) to bivalent regions in mouse ES cells [53]. Interestingly, we also identified the histone lysine 9 and 36 demethylase KDM4C/JMJD2C [54,55] that colocalises with EZH2 in

mouse ES cells [56], and PHF8, a demethylase involved in removal of H3K9me2, H3K27me2 and H4K20me1-residues [57,58]. Among the core components of the PRC1 and PRC2 complexes, we also detected BCOR, MGAP and LMBL2 which are part of the alternative PRC1.1 and PRC1.6 complexes [59]. Notably, we also observe the methylcytosine dioxygenase TET1 and the transcriptional repressor SIN3A being associated to bivalent sites, in line with previous genomic studies showing TET1 and SIN3A at bivalent promoters in ES cells [60]. Finally, we introduced the bivalent reader in ES cells lacking H3K27me3 and performed ChromID to control for false-positive proteins stemming from unspecific interactions of the readers with marks outside of nucleosomes modified by H3K27me3 and H3K4me3. In this case we could not identify any significantly enriched proteins, indicating that the reported proteins are indeed localised to bivalently-modified regions in the genome (Supplementary Fig. 13e).

To exclude that the BASU biotin ligase could influence genomic localisation of the readers and therefore falsely report proteins from sites not decorated by the targeted modifications, we performed biotin-ChIP-seq of the eCR-BASU constructs and compared their binding to the previously-obtained datasets of eGFP-fusion constructs (Supplementary Fig. 14a-c). Based on our genome-wide analysis, binding was highly correlated between the same readers fused to either eGFP or BASU, indicating that the reader domains are not influenced by the addition of the biotin ligase. In summary, these results highlight the applicability of modular eCRs as a platform for biotin ligase recruitment, enabling successful identification of the associated proteins of chromatin subtypes.

Integrative analysis of ChromID datasets reveals the chromatin preference of regulatory proteins

Based on the combined datasets from all ChromID experiments we investigated the distribution of proteins between the different chromatin states, revealing proteins shared between multiple chromatin states, and proteins specific to single chromatin modifications. The latter was most prominent for H3K4me3-associated proteins (Fig. 5c, Supplementary Fig. 15a). Notably, several proteins identified at H3K27me3 or bivalent regions were also associated with H3K4me3, which is expected given the overlap of these modifications in ES cells [47,48]. In addition, we found multiple proteins shared between H3K9me3 and H3K27me3, while little overlap was found between the H3K9me3 set and proteins detected by H3K4me3 or bivalent regions (Fig. 5c and Supplementary Fig. 15a). Given the well-established crosstalk between DNA methylation and H3K9me3, we identified a substantial overlap between these sets (Supplementary Fig. 15b). The functional relationship of the detected factors was further visualised from high-confidence interaction scores obtained from the STRING database [61], revealing a strong interconnectivity between proteins and complexes associated with H3K4me3 and bivalent regions or with heterochromatin marked by H3K9me3 or H3K27me3 (Fig. 5d).

To obtain a quantitative view on the distribution of regulatory proteins along the interrogated chromatin marks, we clustered factors from different regulatory groups based on their enrichment across all datasets (Fig. 5e). For transcription factors, we observed several associations with H3K4me3 regions (e.g. SP2, MAX, FOXK2, ZFX). In addition, several TFs, mainly uncharacterised ZNF proteins, are associated with DNA methylation (ZNF280B, ZNF292, ZNF462, CASZ1, TCF20) and we also recover TFs previously identified to interact with methylated DNA in pull-down or HT-SELEX assays (KLF4, RREB1, ZNF191/24) [9,46] (Fig. 5e). Similar to TFs, chromatin remodellers separate into a group predominantly associated with H3K4me3 (e.g. BRD2, BRD4, INO80, CECR2), and a group preferentially associated with closed chromatin (e.g. ATRX, BAZ2A, SMARCA1). Chromatin writers such as H3K4-specific methyltransferases (KMT2A, KMT2B) and histone acetyltransferases (EP300) were preferentially located at H3K4me3, while writers of repressive marks associated with DNA

methylation, H3K9me3 and/or H3K27me3 (DNMT1, EHMT1/2, NSD1, EZH2). Furthermore, and in line with genome-wide binding data, erasers such as KDM2A, KDM2B, TET1 or KDM5A were predominantly found at H3K4me3 sites, and we find several DNA repair factors associated with repressive chromatin marks (Fig. 5e). Taken together, these datasets obtained by ChromID provide a valuable resource of chromatin-mediated protein interactions in the ES cell genome.

Discussion

Here we present ChromID, a quantitative approach that enables identification of proteins associated with individual and combinatorial chromatin modifications in living cells. ChromID takes advantage of the affinity of engineered chromatin readers, which we obtained from natural reader domains of well-characterised chromatin regulators (CBX1, CBX7, dPC, TAF3, MBD1 and MeCP2). First, we characterised and functionally validated the binding selectivity of all eCRs using a series of quantitative and functional methods in mouse ES cells. The obtained results highlight the applicability of eCRs as an alternative to antibodies for studying subnuclear localisation, genome-wide distribution and histone-PTM combinations in living cells. Interestingly, single domain eCRs were often insufficient to achieve binding to histone modifications under physiological conditions. This is in line with several well-known examples where binding of full-length proteins or complexes to chromatin rely on multivalent interactions [62-64], including recent studies that introduced synthetic chromatin readers for immunofluorescence or activation of reporter genes [65-67]. We made use of the required multivalent interactions to generate synthetic readers that recognise two modifications on the same nucleosome. This allowed us to directly target genomic sites that are bivalently marked by H3K4me3 and H3K27me3, providing novel tools to study and manipulate chromatin modifications in a context-dependent manner.

Finally, to identify the chromatin-associated proteome, we developed ChromID where we use the eCRs to tether promiscuous biotin ligases to chromatin, resulting in biotinylation of proteins in a ~ 35 nm radius around the modification of interest. This allowed us to detect proteins that directly and indirectly associate with chromatin modifications including DNA methylation, H3K4me3, H3K27me3 and H3K9me3, resulting in a total of 518 identified proteins. Among these we identified 180 high-confidence proteins enriched across all datasets, enabling us to assign factors based on their preference towards single or multiple chromatin marks. By employing the bivalent reader, we further achieved specific identification

388 of proteins bound at sites marked by H3K4me3 and H3K27me3, revealing the presence of
389 activating and repressing proteins from Trithorax and Polycomb complexes and additional
390 factors that could play a role in chromatin regulation at bivalent sites.
391
392 Overall, the results from the individual and combinatorial measurements highlight
393 ChromID as a novel approach to uncover how protein recruitment is influenced by chromatin
394 modifications in living cells. Importantly, the usage of natural reader domains in ChromID
395 mimics physiological engagement of proteins with chromatin. This has several benefits, since
396 the eCR-mediated interactions do not require crosslinking or single-stranded DNA. Another
397 benefit of ChromID is the usage of proximity biotin ligation to label and subsequently identify
398 the proteins associated with different chromatin flavours in a unified manner. This enables
399 comparative studies between different chromatin modifications, circumventing the necessity
400 of antibodies, which have always been limiting in such assays due to their variation in affinity
401 and avidity, lack of availability and cost. Furthermore, once biotinylated, the proteins are
402 enriched using highly-stringent washing and elution conditions, ensuring effective removal of
403 background signals and reproducible detection and multi-sample comparison.
404
405 We expect ChromID to be used to chart the genome-wide proteome at multiple
406 chromatin modifications and in numerous cell types in order to understand how the chemical
407 language on chromatin directs protein recruitment in a spatiotemporal manner. The
408 applicability of ChromID in living cells, as well as eCRs as synthetic readers, further opens the
409 possibility to perform similar experiments in a tissue-specific manner in living animals,
410 providing exciting future avenues to chart the epi-proteome during dynamic cellular processes
411 and development.

Acknowledgements

We thank Dirk Schuebeler (FMI, Basel) for providing the *Dnmt*-TKO ES cell line and Paul A. Khavari (Stanford) for cDNA encoding the BASU biotin ligase. Furthermore, Bernd Roschitzki, Jonas Grossmann and Tobias Kockmann (FGCZ), Philip Knobel (USZ) and Mario Majchrzak and Robin Klemm (UZH) for the initial discussions on biotin-ID methods and detction. We would like to thank members of the Functional Genomics Center Zurich for high-throughput sequencing and mass spectrometry support, José María Mateos and Joana Delgado Martins from the Centre for Microscopy and Image Analysis, and the Science IT (S3IT) team at the University of Zurich for providing the computational infrastructure. We are grateful to the Edinburgh Protein Production Facility (EPPF) for their support. We thank Viktoria Major for help with cloning and Kimberly Webb for help with histone purification. Furthermore, we thank Matthias Altmeyer (UZH), Arnaud Krebs (EMBL), Dirk Schuebeler (FMI), and members of the Baubec Lab for their critical input on the manuscript. Research in the TB Lab is supported by the Swiss National Science Foundation (SNF Professorship #157488 and SNF Sinergia #180345), the Swiss initiative in Systems Biology (SystemsX.ch), the Alumni Association of UZH (ZUNIV/FAN), and the University of Zurich. NS and CF acknowledge individual support by EMBO long-term fellowships. Work in the RA laboratory is supported by ERC (AdvGr 670821 (Proteomics 4D)) and Innovative Medicines Initiative (EU/EFPIA) [ULTRA-DD grant no. 115766]. Research in the PV lab is supported by the Wellcome Trust [104175/Z/14/Z, Sir Henry Dale Fellowship to P.V.] and ERC (ERC-STG 639253). EB was supported by the Wellcome Trust through a Doctoral Studentship [105244]. The Wellcome Centre for Cell Biology was supported by core funding from the Wellcome Trust [203149]. The EPPF was supported by the Wellcome Trust [101527/Z/13/Z].

Author Contributions

RV and TB conceived and designed the study. RV, RP, NS, MM and TB developed the tools and protocols. RV, RP, SB, SG, CA and JW generated cell lines and performed experiments. C.F. and R.A. designed and performed LC–MS experiments, analysed data and interpreted results for histone-PTM detection. AG performed STRING network analysis with supervision from CvM. EB and TS performed nucleosome reconstitution and interaction experiments under supervision from PV. RV, RP, MM and TB analysed data. RV and TB wrote the manuscript with input from all authors.

Declaration of Interests

The authors declare no competing interests.

Data access information for reviewers

GEO accession GSE128907:

Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128907

Enter token uxchscsuptkfhsd into the box

PRIDE accession PXD014483: https://www.ebi.ac.uk/pride/archive/

Username: reviewer12463@ebi.ac.uk

Password: L5Udr6Y1

Methods

Molecular cloning:

DNA encoding for wild type and mutant domains were amplified from cDNA or synthesized (IDT technologies) based on available domain annotations (Uniprot). These sequences are introduced in-frame into the EcoRV site of the RMCE-targeting vector parbit-v6 by Gibson assembly. The final construct expresses the N-terminal biotin-tagged domain of interest fused in-frame to a cassette containing an NLS signal followed by eGFP, an internal ribosome entry site (IRES) and the puromycin-N-acetyltransferase gene (pac), which confers resistance against puromycin treatment. All coding sequences are under control of a constitutive CAG promoter. BioID2-HA and the 13X-Linker were PCR amplified from MCS-13X-Linker-BioID2-HA plasmid (Addgene, plasmid #80899), HA-BASU was PCR amplified from BASU-RaPID plasmid (kindly provided by Paul Khavari; equivalent to Addgene plasmid #107250). PCR-amplified products were cloned into RMCE-targeting vector L1-CAG-NLS-IRES-pac-1L (parbit-v9) using Gibson assembly. For bacterial expression of eCR-eGFP-6xHis fusion proteins, sequences spanning the domains of interest along with the NLS and eGFP were PCR amplified from parbit-v6 and subcloned into a modified pET-28 vector encoding an in-frame C-terminal 6xHis affinity tag.

Cell culture and cell line generation:

Mouse embryonic stem cells (129×C57BL/6) were cultured on 0.2% gelatin-coated dishes in ES medium containing DMEM (Invitrogen;), supplemented with 15% fetal bovine serum (FBS; Invitrogen), 1x non-essential amino acids (Invitrogen), 2mM L-glutamine (Invitrogen), 0.01% 2-mercaptoethanol (Sigma Aldrich), and homemade leukaemia inhibitory factor (LIF) at 37 °C in 7% CO2. Cell lines were obtained by recombinase-mediated cassette exchange (RMCE). Briefly, engineered chromatin-reader containing RMCE constructs were co-transfected with a CRE recombinase expression plasmid (1: 0.6 µg DNA ratio) to RMCE-competent and biotin ligase (BirA)-positive mouse ES cell lines [22]. Two rounds of selection were applied to yield a

homogenous population of eCR expressing cells: i) cells that did not exchange the RMCE-cassettes were first removed by negative selection (3mM ganciclovir for 4 days), ii) cells expressing the eCR of interest were retained by positive selection (2mM puromycin for 2 days). Homogenous and stable protein expression was then monitored by measuring eGFP expression using flow cytometry, immunofluorescence (IF) and immunoblotting. Transfections were carried out using Lipofectamine 3000 reagent (Thermo Fisher Scientific, L3000015) at a 2:1 µg DNA ratio in OptiMEM medium (Thermo Fisher Scientific, 31985070). The *Eed*-KO cell line was generated by transfecting px330 with a guide (GGTGAAAAAATAATGTCCTG) targeting exon 8 together with a recombination-repoter [68]. 36 hours after transfection, cells were treated with 2 ug/ml puromycin for 36 h. Single ES cell colonies were harvested and gDNA was screened by PCR (forward primer: GAGGCACTAGGTGAAAACTGGA, reverse primer: CACAGAGCCTGTCACCTGTT). Positive KO clones were validated by Sanger sequencing (Microsynth) and H3K27me3 Western blotting. The endogenously tagged *Zfp280D* cell line was generated with a guide (AGTAGACCTGGCAGATGGAG) targeting exon 22, using the same approach outlined above. 72 hours after transfection, single GFP-positive cells were sorted into 96-wells by flow cytometry. Single ES cell colonies were harvested and gDNA was screened by PCR (forward primer: AGCACAAATCCCACTGAAGCTG, reverse primer: TCGTGCCATTCGATTTTCTGAGC). Positive clones were validated by Sanger sequencing. Neuronal differentiation of ES cells was performed as previously described [69].

Flow cytometry:

Cells were harvested by trypsinisation and resuspended in DPBS for analysis. If necessary, cells were incubated with LIVE/DEADTM Fixable Near-IR Dead Cell Stain (Invitrogen, L34975) to discriminate cell viability. Samples were analysed for eGFP expression and LIVE/DEADTM stain by flow cytometry on a FACSCanto (BD Biosciences). Cells were gated for viable and individual cells, channel voltages for eGFP (Alexa Fluor 488-A) and live/dead (APC-Cy7A) signals were set regarding verified negative and positive eGFP-expressing control cells. Raw files were analysed and visualised using FlowJo software (Tree Star;

version 10.0.7). For CD24 measurements in neuronal progenitors, single cell suspensions were obtained from neuronal progenitors after 8 days of differentiation through trypsinization and filtered through 40-um cell strainers (BD Biosciences), as previously described [69]. For cell surface staining cells were incubated for 30 min at 4°C with saturating concentration of anti-CD24a monoclonal antibody in the presence of anti-CD16/CD32 (eBioscience). For live dead cell exclusion LIVE/DEAD Fixable Near-IR Dead Cell Stain (Invitrogen) was used. Samples were acquired using FACSFortessa (BD Biosciences) and data were analysed using FlowJo software (Tree Star).

Western blotting:

For western blotting, 20 µg of protein were resolved in NuPAGE-Novex Bis-Tris 4–12% gradient gels (Invitrogen) and transferred on polyvinylidene fluoride (PVDF) membranes in a tank containing transfer buffer (50mM Tris, 0.4M glycine, 0.05% SDS, and 20% MeOH). The membrane was washed once with TBST (10mM Tris pH 8.0, 150mM NaCl, and 0.1% Tween-20), blocked with 5% (w/v) BSA in TBST, and stained with the corresponding primary antibody anti-HP1b/CBX1 (1:1,000, CST; #8676) or Lamin B1 (1:1000, Santa Cruz, sc-374015) at 4°C overnight. After washing three times with TBST for 10 minutes, the signal was detected with corresponding species-specific horseradish peroxidase (HRP)-conjugated secondary antibodies. Three additional TBST washes were applied prior protein detection using the Amersham ECL Western blotting detection reagent (GE Healthcare Life Sciences; RPN2109) with subsequent signal development on Amersham Hyperfilm ECL (GE Healthcare Life Sciences; 28906836). For validation of Eed-KO cell lines, cells were lysed with NETN buffer (20 mM Tris (pH 8), 0.5 % (v/v) NP-40, 100 mM NaCl, 1 mM EDTA (pH 8)) supplemented with 1 x protease inhibitor cocktail (Roche; COEDTAF-RO) and 1 mM DTT (Sigma Aldrich, DTT-RO). Nuclei were pelleted at 6,500 x g for 10 min at 4 °C, washed once with NETN. Histones were acid-extracted overnight at 4 °C in 0.2 N HCl at a density of 4 x 107 nuclei per ml. Histone extracts were then centrifuged at 6,500 x g for 10 min at 4 °C to pellet debris and 5 ug were loaded onto a NuPAGE-Novex 16% Tris-Glycine Gel (Invitrogen). Western Blot and protein

542 detection were performed as above with a transfer buffer containing no SDS, but 20 % MeOH,
543 and membrane was stained with primary antibody anti-H3K27me3 (Diagenode, C15410195),
544 anti-Histone H1 (Millipore, 05-457) and anti-Histone H3 (Abcam, ab1791).
545
546 Live-cell imaging and image processing:
547 2x 10^4 eCR-eGFP fusion expressing cells were seeded on 0.2% gelatin-coated 35-mm glass
548 bottom chambers (Ibidi; 80826) one day before imaging. Next day, cells were stained with
549 Hoechst 33342 (Invitrogen; 62249) for 10 minutes, washed twice with DPBS, and covered
550 with ES cell imaging medium containing DMEM (Invitrogen; 31053028), supplemented with
551 15% foetal bovine serum (FBS; Invitrogen), 1x non-essential amino acids (Invitrogen), 2mM
552 L-glutamine (Invitrogen), 0.01% 2-mercaptoethanol (Sigma Aldrich), and homemade
553 leukaemia inhibitory factor (LIF). Randomly selected cells were imaged with sequential
554 acquisition settings on a Leica SP5 inverted confocal laser scanning microscope equipped
555 with a climate chamber, an Argon laser for 453, 476, 488, 496, and 514 nm, and a diode laser
556 for 561 nm. The filters for fluorescence imaging were GFP (ex BP 470/40, em BP 525/50) and
557 N3 (ex BP 546/12, 600/400). Confocal images were acquired with an HCX PL APO Leica 63×
558 oil immersion objective with HyD detectors. Z stacks were acquired per site using a 0.3 μm
559 step size.
560 Time-lapse fluorescence microscopy was performed with a confocal spinning disk imaging
561 system (Olympus IXplore SpinSR10, Olympus Corporation) equipped with a CSU-W1 unit
562 (YOKOGAWA) and a 60× UPLSAPO UPlan S Apo silicon oil objective of 1.3 NA (Olympus
563 Corporation). 11 z planes were acquired per site (1μm step size) every 5 min for approximately
564 12.5 hours. A 488nm laser was used to excite the GFP probe while emitted light was filtered
565 by a 525/50 band pass filter and captured by a Prime BSI Scientific CMOS camera (2048 ×
566 2048 pixels, Teledyne Photometrics). Images were deconvolved using Huygens Professional
567 19.10 software (Scientific Volume Imaging) using up to 40 iterations of the Classic Maximum
568 Likelihood Estimation algorithm with a theoretical PSF. Background correction was automatic.

The signal-to-noise ratio setting was adjusted empirically to 16 to give satisfactory results. Image analysis was performed on the resulting image series using FIJI (version 2.0.0) and the Bio-Formats Importer plugin. Appropriate single z-planes were then selected for further image analysis and display.

Immunostaining and immunofluorescence:

Cells were grown in 0.2% gelatine-coated 35-mm glass bottom chambers (Ibidi; 80826), fixed in 4% formaldehyde in PBS for 10 min at room temperature, washed three times in PBS, permeabilized for 5 min at room temperature in PBS supplemented with 0.1% Triton X-100 and 0.25% BSA (Sigma-Aldrich), and washed twice in PBS. Corresponding primary (anti-H3K9me3: ab8898 (abcam), anti-H3K27me3: C15410195 (diagenode), anti-GFP: 11814460001 (Millipore) and secondary antibodies (Alexa Fluor 488 anti-mouse and 568 anti-rabbit IgGs from ThermoFisher) were diluted in PBS containing 2% FBS and 0.02% BSA. Primary antibody incubations were performed overnight at 4°C. Secondary antibody incubations were performed for 1h at room temperature. Following antibody incubations, cells were washed once with PBS and incubated for 10 min with PBS containing 4',6-diamidino-2-phenylindole dihydrochloride (DAPI, 0.5 µg/ml) for 10 minutes at room temperature to stain DNA. Following three washing steps in PBS, cells in multi-well plates were kept in PBS for imaging. Randomly selected cells were imaged with sequential acquisition settings on a Leica SP5 inverted confocal laser scanning microscope.

Generation of recombinant nucleosomes:

Xenopus H3 and H4 and human H2A and H2B were expressed from pET-3a or pET-3d vectors in BL21 (DE3) pLysS for H3, H2A, and H2B or BL21 (DE3) for H4 through induction with 0.2 mM IPTG for 4 h at 37°C. Histones were purified from inclusion bodies and solubilised in unfolding buffer (20 mM Tris-HCl pH 7.5, 7 M guanidine HCl, 10 mM DTT). Extracted histones were dialysed against three changes of urea dialysis buffer (10 mM Tris HCl pH 8,

7 M urea, 100 mM NaCl, 1 mM EDTA, 5 mM beta-mercaptoethanol, this and all subsequent histone dialysis steps were carried out at 4°C) and then purified further by passing over a HiTrap Q column (GE Healthcare) before binding to and NaCl gradient elution from HiTrap SP cation exchange chromatography columns (GE Healthcare). Fractions containing histones were pooled and dilaysed three times against water containing 5 mM beta-mercaptoethanol and lyophilised for long-term storage at -80°C. To express histones for native chemical ligation (NCL), constructs encoding truncated *Xenopus* histone H3 were generated in pET-3a. For generation of H3K4me3-modified histones, truncated H3 lacking residues 1-31 after the initiator methionine, with a threonine-to-cysteine substitution at position 32 of *Xenopus* H3 and a cysteine-to-alanine substitution at position 110 (H3Δ1–31 MT32C C110A) was expressed in BL21 (DE3) pLysS and purified as above, except for the final dialysis, which was carried out as two rounds of dialysis against 1 mM DTT in H_2O and one round against 0.5 mM TCEP before lyophilisation and storage. For generation of H3K27me3-modified histones, a similarly truncated Xenopus H3 construct was used, lacking the first 44 residues and carrying a threonine-to-cysteine mutation at residue 45 (H3Δ1–45 MT45C C110A).

Native chemical ligation reactions were carried out in 6 M Guanidine HCl, 250 mM sodium phosphate buffer pH 7.2, 150 mM 4-mercaptophenylacetic acid (MPAA), 50 mM TCEP for 72 h at room temperature with constant agitation. Reactions were then dialysed three times against urea dialysis buffer (see above, but with 1 mM DTT instead of 5 mM beta-mercaptoethanol). Ligated full-length modified histones were separated from unligated histone through cation exchange chromatography on a HiTrap SP column (GE) and then dialysed against three changes of water containing 5 mM beta-mercaptoethanol before lyophilization and storage at -80°C until use. For H3K4me3 and H3K4me1-modified histones, H3Δ1–31 MT32C C110A was reacted with a synthetic peptide spanning residues 1–31 of histone H3.1 containing tri- or mono-methylated lysine at position 4 and a C-terminal benzyl thioester (Peptide Protein Research Ltd., Fareham, UK). For H3K27me3-modified histones, H3Δ1–44

622 MT45C C110A was reacted with a synthetic peptide spanning H3.1 residues 1–44 including
623 trimethylated lysine at position 27 and a C-terminal benzyl thioester.
624 To reconstitute histone octamers, the four core histones were resuspended in unfolding buffer
625 (see above), mixed in a mass ratio of 1:1:1.2:1.2 (H3:H4:H2A:H2B), and dialysed against three
626 changes of refolding buffer (10 mM Tris HCl pH 8, 2 M NaCl, 1 mM EDTA, 5 mM beta-
627 mercaptoethanol) at 4°C. After centrifugation to remove precipitate formed during dialysis,
628 correctly assembled histone octamers were purified by size exclusion chromatography in
629 refolding buffer on a S200 column (GE Healthcare) using an Akta PURE system (GE
630 Healthcare). DNA template for mononucleosome assembly was generated by PCR with a
631 biotinylated forward primer, amplifying a 209-bp fragment centered around the 147-bp 601
632 nucleosome positioning sequence, followed by PCR purification and elution into TE buffer. To
633 reconstitute recombinant mononucleosomes, DNA and histone octamers were combined in
634 refolding buffer supplemented with 5 M NaCl to compensate for reduction in NaCl
635 concentration due to introduction of TE buffer with the DNA, followed by gradient dialysis
636 against TE buffer down to 400 mM NaCl and then a step dialysis against TE buffer at 4°C.
637 Optimal ratios of DNA and histone octamer were determined so that at least 95% of DNA was
638 complexed, but without over-assembly and unspecific DNA binding of histones. Assemblies
639 were routinely checked by native gel electrophoresis on 6% acrylamide gels in TGE buffer.
640
641 Bacterial expression of eCR-eGFP-His fusion proteins:
642 eCR-eGFP fusion proteins were expressed in BL21 (DE3) *E. coli* by induction for 3 h at 37°C
643 with 0.5 mM IPTG in the presence of 20 µM $ZnCl_2$. Cells were resuspended in lysis buffer
644 (20 mM Tris HCl pH 8, 500 mM NaCl, 0.1% NP-40, 0.5 mM PMSF) and lysed by sonication.
645 Lysates were cleared by centrifugation at 23,000 g for 30 min at 4°C. Cleared lysates were
646 incubated with Sepharose 6 Fast Flow Ni-NTA resin (GE Healthcare) for 1 h. All purification
647 steps were carried out at 4°C. Beads for collected by brief centrifugation at 800 g,
648 resuspended in 10 ml of 300 mM wash buffer (50 mM NaH_2PO_4 pH 8, 300 mM NaCl, 20 mM
649 imidazole, 0.1 mM PMSF), and transferred to a 10-ml polyprep column (Biorad). Settled beads

were then washed once with 10 ml of 1 M wash buffer (see above, but with 1 M instead of 300 mM NAcl), and once with 5 ml of wash buffer (see above, but without PMSF), before elution with 5 ml elution buffer (50 mM NaH_2PO_4 pH 8, 300 mM NaCl, 250 mM imidazole), collecting 0.5-ml fractions. Fractions were analysed by Bradford protein assay (Biorad) and SDS-PAGE. Fractions containing the desired eCR-eGFP fusion protein were pooled and dialysed against three changes of BC100 (20 mM HEPES KOH pH 8, 100 mM KCl, 10% glycerol, 0.5 mM DTT) at 4°C and re-analysed by Bradford and SDS-PAGE.

Nucleosome pulldown assays:

For pulldown assays with recombinant modified nucleosomes and eCR-eGFP fusion proteins, streptavidin sepharose high performance beads (GE Healthcare, 8 µl of slurry per pulldown) were briefly washed three times with pulldown buffer (20 mM HEPES KOH pH 7.9, 150 mM NaCl, 10% glycerol, 1 mM EDTA, 1 mM DTT, 0.2 mM PMSF, 0.1% NP-40). All centrifugation steps were carried out at 1,500 g for 2 min at 4°C. All incubation steps were carried out at 4°C. After washes, beads were incubated overnight with 3 µg of assembled recombinant nucleosomes (in TE diluted with pulldown buffer and adjusted to 0.1% NP-40 and 150 mM NaCl final concentration by addition of 10% NP-40 and 5 M NaCl, respectively) with constant rotation in the cold room. Beads were then collected by centrifugation and washed briefly with three changes of pulldown buffer. Bead-bound nucleosomes were then incubated with increasing amounts of eCR-eGFP fusion proteins for 2 h under constant agitation in the cold room. Beads were then washed with pulldown buffer by 5-min incubation under rotation. For TAF3 Phd eCR fusions, two washes with pulldown buffer containing NP-40 followed by three washes in pulldown buffer without NP-40 were performed. For CBX7 Chromo eCR fusions, four washes in pulldown buffer with NP-40 followed by two washes without NP-40 were carried out. After washes, bound proteins were eluted from beads by boiling for 5 min at 95°C with 1.5x SDS sample buffer (95 mM Tris HCl pH 6.8, 15% glycerol, 3% SDS, 75 mM DTT, 0.15% bromophenol blue). Protein binding was analysed by Western Blotting with anti-His antibody (Sigma, H10229), anti-H4 antibody (Cell Signaling Technology, 13919), anti-

H3K4me3 (abcam, ab8580), and anti-H3K27me3 (Cell Signaling Technology, 9733). 40% of bound sample was loaded along with 20% of input samples.

Biotin chromatin immunoprecipitation (bio-ChIP):

For cross-linking and chromatin extraction, 30–50 × 10^6 cells were harvested by trypsinisation, washed once, and fixed for 8 min with 1% formaldehyde at room temperature in medium containing DMEM (Invitrogen), supplemented with 15% foetal bovine serum (FBS; Invitrogen) followed by the addition of glycine (final concentration 0.12 M) and incubation for 10 min on ice. Cells were collected by centrifugation at 680 g for 5 min and washed twice with 10 ml DPBS. After last wash, cells were resuspended in 10 ml buffer containing 0.25% Triton X-100, 1 mM EDTA, 10 mM TRIS, 0.5 mM EGTA, and 200 mM NaCl and incubated for 10 min on ice followed by centrifugation at 680 g for 5 min. Final cell lysis was performed in MNase Digest Buffer (20 mM Tris (pH 8.0), 5 mM MgCl2, 1 mM CaCl2, 10 mM NaCl, 0.25 M Sucrose, 1 % Triton-X100 and 1x EDTA-free complete protease inhibitor cocktail (PIC, Roche; COEDTAF-RO)) in a concentration-dependent volume (2x 10^4 cells per µl) for 1h on ice. For Mnase digestion, cross-linked chromatin was incubated at 37 °C for 2 minutes followed by addition of 50U Mnase per ml and additional incubation at 37 °C for 20 minutes. Digestion was stopped by addition of stop-buffer (150mM NaCl, 100 mM EDTA, 200 mM EGTA) followed by addition of SDS to a final concentration of 0.1% and 4 cycles (32/50 seconds ON/OFF) sonication in a Bioruptor Pico instrument (Diagenode) according to the manufacturer's instructions. Fragmented chromatin was centrifuged at 12,000 g for 10 min at 4°C and supernatant was used for further steps. Streptavidin-M280 magnetic beads were blocked for 1 h with 1% cold fish skin gelatine (Sigma Aldrich) and 100 ng tRNA (Sigma Aldrich) supplemented with protease inhibitor cocktail mix (Roche) and washed twice with buffer 3 with 0.1% SDS and 150 mM NaCl. Chromatin (150–200 µg) was then incubated with 90 µl pre-blocked streptavidin-M280 magnetic beads overnight at 4°C. Beads were washed under rotation for 8 min for each wash step and placed on a magnetic rack for 2 min for exchange

of buffers first with two rounds of 2% SDS, high salt buffer (50 mM HEPES pH 7.5 (adjusted with KOH), 500 mM NaCl, 1mM EDTA, 1 % Triton-X100, 0.1 % deoxycholate, 0.1% SDS), DOC buffer (250 mM LiCl, 0.5% NP-40, 0.5% deoxycholate, 1 mM EDTA, 10 mM TRIS), and two rounds of Tris/EDTA buffer. Beads were treated with RNaseA (60 µg, Roche) for 30 min at 37°C in 1% SDS, 0.1 M NaHCO3, and subsequently proteinase K (60 µg, Roche) for 3 h at 55°C in 1% SDS, 0.1 M NaHCO3, 10 mM EDTA, 20 mM TRIS, followed by de-cross-linking overnight at 65°C. DNA was purified with phenol–chloroform extraction and ethanol precipitation.

Antibody chromatin immunoprecipitation (AbChIP):

For cross-linking and chromatin extraction, 30–50 × 10^6 cells were harvested by trypsinisation, washed once, and fixed for 8 min with 1% formaldehyde at room temperature in medium containing DMEM (Invitrogen), supplemented with 15% foetal bovine serum (FBS; Invitrogen) followed by the addition of glycine (final concentration 0.12 M) and incubation for 10 min on ice. Cells were collected by centrifugation at 680 g for 5 min and washed twice with 10 ml DPBS. After last wash, cells were resuspended in 10 ml buffer containing 0.25% Triton X-100, 1 mM EDTA, 10 mM TRIS, 0.5 mM EGTA, and 200 mM NaCl and incubated for 10 min on ice followed by centrifugation at 680 g for 5 min. Final cell lysis was performed in Mnase Digest Buffer (20 mM Tris (pH 8.0), 5 mM MgCl2, 1 mM CaCl2, 10 mM NaCl, 0.25 M Sucrose, 1 % Triton-X100 and 1x EDTA-free complete protease inhibitor cocktail (PIC, Roche; COEDTAF-RO) in a concentration-dependent volume (2x 10^4 cells per µl) for 1h on ice. For Mnase digestion, cross-linked chromatin was incubated at 37 °C for 2 minutes followed by addition of 50U Mnase per ml and additional incubation at 37 °C for 20 minutes. Digestion was stopped by addition of stop-buffer (150mM NaCl, 100 mM EDTA, 200 mM EGTA) followed by addition of SDS to a final concentration of 0.1% and 4 cycles (32/50 seconds ON/OFF) sonication in a Bioruptor Pico instrument (Diagenode) according to the manufacturer's instructions. Fragmented chromatin was centrifuged at 12,000 g for 10 min at 4°C and

supernatant was used for further steps. 50 ul per IP of protein-A/G magnetic beads (Invitrogen) were blocked for 1 h at 4 °C with 100 ng tRNA (Sigma Aldrich) and 100 ng BSA (Sigma Aldrich) supplemented with 1 x protease inhibitor cocktail mix (PIC, Roche), washed twice with 1 x TE and taken up in 50 ul/IP 1 x TE and 1 x PIC. 100 µg of Chromatin were then pre-cleared with 20 µl pre-blocked protein A/G magnetic beads for an hour at 4 °C. After bead removal, 5 ug/IP of antibody were added (H3K27me3: Diagenode, C15410195, H3K4me3: abcam, ab8580, H3K9me3: abcam, ab8898) at 4 °C under over-night rotation. 30 ul of blocked beads were added for another 3-4 hours of immunoprecipitation. Beads were washed at 4 °C under rotation for 8 min for each wash step and placed on a magnetic rack for 2 min for exchange of buffers with two rounds of high salt buffer (50 mM HEPES pH 7.5 (adjusted with KOH), 500 mM NaCl, 1mM EDTA, 1 % Triton-X100, 0.1 % deoxycholate, 0.1% SDS), one round of DOC buffer (250 mM LiCl, 0.5% NP-40, 0.5% deoxycholate, 1 mM EDTA, 10 mM TRIS), and two rounds of Tris/EDTA buffer. Beads were treated with RNaseA (60 µg, Roche, RNASEA-RO) for 30 min at 37°C in 1% SDS, 0.1 M NaHCO3, and subsequently proteinase K (60 µg, Roche) for 3 h at 55°C in 1% SDS, 0.1 M NaHCO3, 10 mM EDTA, 20 mM TRIS, followed by de-crosslinking overnight at 65°C. DNA was purified with Qiagen MinElute Kit according to manufacturer's instructions and eluted in 10-20 ul elution buffer.

ChIP-seq library preparation, high-throughput sequencing and read pre-processing:

For ChIP-seq, libraries were prepared using the NEB-next ChIP-seq library Kit (E6240) following the standard protocols. Up to 8 samples with different index barcodes were combined at equal molar ratios and sequenced as pools. Sequencing of library pools was performed on Illumina HiSeq 4000 or Nova Seq machines according to Illumina standards, with 75- to 150-bp single-end sequencing. Library demultiplexing was performed following Illumina standards. Samples were filtered for low-quality reads and adaptor sequences removed using Trim Galore (https://github.com/FelixKrueger/TrimGalore). Filtered reads were mapped to the mouse genome (version mm9) using the BOWTIE algorithm allowing for two

759 mismatches and only unique mappers were used (-m 1 --best --strata). Identical reads from
760 PCR duplicates were filtered out.
761
762 Selection of genomic coordinates and data analysis:
763 To obtain genome-wide 1kb intervals, we partitioned the entire genome into 1 kb sized tiles.
764 Intervals overlapping with satellite repeats (Repeatmasker), ENCODE black-listed and low
765 mappability scores [70] (below 0.5) were removed in order to reduce artefacts due to annotation
766 errors and repetitiveness. To detect eCR-enriched regions, we utilised MACS2 using the
767 eGFP ChIP-seq as background and applying the following parameters: --broad -g mm --broad-
768 cutoff 0.1. To detect antibody-specific peaks for histone modifications, we applied the same
769 approach but using input chromatin as a background signal. Obtained histone modification
770 peaks were further filtered according to qval >= 2 and pileup >= 3.4 scores.
771 Peaks were overlapped with genomic features and coverages were calculated using the
772 following hierarchy: promoters, enhancers, exons, repeats and introns. Promoters were
773 defined as +/- 1kb around RefSeq gene TSS, enhancers were defined based on DHS peaks
774 where H3K4me1 was higher than H3K4me3 [71], exons and introns were retrieved based on
775 RefSeq annotations, and repetitive elements using Repeatmasker. ChromHMM segmentation
776 [72] of the mouse genome was obtained from http://compbio.mit.edu/ChromHMM/ , as part of
777 ENCODE.
778 For Figure 1e, a genomic range object containing all peaks was generated, overlapping peak
779 regions were merged, and finally used to compute correlations between eCR and antibody
780 signals. To define H3K4me3-monovalent, H3K27me3-monovalent and bivalently-marked
781 peaks, we calculated the H3K4me3 and H3K27me3 enrichments at these sites and selected
782 all H3K4me3 peaks devoid of H3K27me3 as H3K4me3-monovalent peaks. H3K27me3
783 lacking H3K4me3 signals were selected as H3K27me3-monovalwent while H3K4me3 peaks
784 positive for H3K27me3 were selected as bivalent peaks.
785 ChIP enrichments at genomic segments and peaks were calculated as $\log_2$-fold changes over
786 input chromatin (for antibodies) and over eGFP (for eCRs) after library size normalization and

using a constant of eight pseudo counts to reduce sampling noise. Heatmap and average density-profiles around peaks were generated using genomation in R (Akalin et al, 2014).

Mass spectrometry analysis of histone modifications:

Histones were processed as described in Feller et al. 2015 with modifications as described below. De-crosslinked histones were separated by SDS-PAGE on 16 % Novex Tris-Glycine gels (Invitrogen, XP00165BOX), stained with InstantBlue (Expedeon, ISB1L) and bands corresponding to core histones were excised. Gel pieces were washed twice with water and twice with 100 mM ammonium bicarbonate. Gel pieces were destained by incubating three times for each 10 minutes with 50 mM ammonium bicarbonate/50% acetonitrile at 37 °C with shaking at 800 rpm in a Thermomixer. Gel pieces were successively dehydrated by incubating with 100 mM ammonium bicarbonate, once 20 mM ammonium bicarbonate and three times with acetonitrile. Histones were twice derivatized by chemical acetylation by reacting 5 µL of d6-acetic anhydride ((CD3CO)2O, Sigma-Aldrich), 15 µL of 100 mM ammonium bicarbonate and 1 M ammonium bicarbonate buffered with 1:2 diluted ammonium hydroxide solution to keep the pH at 8. The reactions were performed for 45 minutes at 37 °C with shaking at 800 rpm in a Thermomixer. After the derivatization reactions, histones were washed four times with 100 mM ammonium bicarbonate, two times with water, and three times with acetonitrile. Histone gel pieces were rehydrated with a 25 ng/ul trypsin solution in 100 mM ammonium bicarbonate (sequencing-grade trypsin from Promega) and digested overnight at 37 °C. Tryptic peptides were extracted twice with 70% acetonitrile/0.25% TFA and twice with acetonitrile, vacuum concentrated in a speed vacuum centrifuge and resuspended in 100 mM ammonium bicarbonate. Free peptide N-termini were derivatized twice by reacting 5 µL of d6-acetic anhydride, 30 µL of 1 M ammonium bicarbonate and 10-15 µL of 1:2 diluted ammonium hydroxide solution to keep the pH at 8. After the reaction mixture was vortexed and incubated at 37 °C for 45 minutes with shaking at 800 rpm in a Thermomixer, it was quenched with 35 µL of 1 M ammonium bicarbonate, the solution volume was reduced to 15 µL in a speed vacuum centrifuge and the derivatization procedure was repeated. After derivatization,

peptides were evaporated in a speed vacuum centrifuge at 37 °C to near dryness, resuspended in 20 µL of 0.1% trifluoroacetic acid (TFA), the pH was adjusted with 50% TFA solution to pH 2 and the peptides were purified by a StageTip protocol using two discs of C18 followed by one disc of activated carbon (3 M Empore). After StageTip purification, the samples were evaporated in a speed vacuum centrifuge to near dryness and stored until mass spectrometry acquisition. Before MS analysis, the histone peptides were resuspended in 20 µL of 0.1% TFA and diluted 1:10 in 0.1% TFA supplemented with iRT reference peptides (Biognosis).

The MS-ready histone solution was separated on a reverse phase liquid chromatography column (Waters Acquity HSS T3, C18, 1.8 µM). The column was connected to a nanoflow UPLC (Waters M-Class) and peptides were electrosprayed in a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). Buffer A was composed of 0.1% formic acid in HPLC-grade water and buffer B was 0.1% formic acid in acetonitrile. Peptides were eluted in a linear gradient with a flow rate of 300 nL per minute, starting at 1% B and ramping to 35% in 55 min, followed by an increase to 45% B in 5 min, followed by an increase to 65% in 3 min followed by an increase to 98% in 5 min and then holding at 98% B for another 10 min. The mass spectrometer was operated in data-independent acquisition mode, ion chromatograms were extracted with Skyline and data summarization and statistical analysis was performed in R and Excel (Feller and Aebersold, manuscript in preparation). In brief, a spectral-chromatogram library was constructed using isotopically-labelled reference synthetic peptides and manually curated spectra from cellular histone preparations. To identify and quantify histone modifications, Thermo raw files were imported into the Skyline histone template and peak identification and ion chromatogram extraction was performed using the spectral-chromatogram library. All peak integrations were manually validated. Relative abundances of acetylated peptidoforms were calculated using the extracted ion chromatograms for all peptides that share the same sequence, including the unmodified peptide, according to the formulas described previously [29].

Nuclear extraction for ChromID:

Cells were cultured with ES medium and induced for the corresponding time periods (12-24 hours) with 50µM biotin (Sigma) dissolved in DPBS. Cells were grown to about 90% confluency on 15cm dishes (approximately 50 x 10^6 cells), harvested by trypsinisation, and pelleted by centrifugation at 1000 rpm for 5min. The subsequent steps were either performed on ice or at 4°C. Pellets were gently resuspended (by shaking) in 5 pellet volumes (PV) of nuclear extract buffer 1 (NEB1; 10mM HEPES pH7.5, 10mM KCl, 1mM EDTA, 1.5mM MgCl2, 1mM dithiothreitol (DTT), and 1x PIC) and swelled on ice for 10 minutes, followed by centrifugation at 2000 g for 10 minutes. Pellets were then gently resuspended in 2x PV of NEB1, followed by dounce homogenisation using a loose pistil (10 times up and down). Nuclei were collected by centrifugation at 2000 g for 10 minutes and resuspended in 1x PV of NEB1 + 12µl/ml of Benzonase (Millipore, 71206) to digest genomic DNA, followed by overhead rotation at 4°C for three hours. Nuclei were then pelleted by centrifugation at 2000 g for 10 minutes, resuspended in 1x PV of NEB2-450 (20mM HEPES pH 7.5, 0.2mM EDTA, 1.5mM MgCl2, 20% glycerol, 450mM NaCl, 1mM DTT, and 1x PIC), Dounce homogenised using a tight pistil (10 times up and down), vortexed, followed by overhead rotation at 4°C for one hour. Cell debris were removed by centrifugation at 2000 g for 10 minutes. The salt concentration of the nuclear extracts (NE) was adjusted to 150mM by drop-wise addition of 2x residual volume of NEB2-NS (see above, without NaCl), and NP40 levels were adjusted to 0.3%. Subsequently, protein concentrations were measured using QubitTM Protein Assay Kit (Thermo Fisher Scientific, Q33211). Equal amounts of proteins were used per IP (standard: 2 mg) and protein lysate volumes were adjusted to equal volumes with IP buffer (IPB; NEB2-150, 0.3% NP40, 1mM DTT, and 1x PIC).

Streptavidin beads preparation for affinity purification:

Streptavidin M-280 Dynabeads (Thermo Fisher) were equilibrated three times with IPB (see above) by overhead rotation for 10 minutes at 4°C and subsequently pre-blocked in IPB + 1% cold fish gelatin rotating 4°C for one hour. Finally, beads were taken up in IPB (same volume

as starting volume). 40 µl of pre-blocked Streptavidin M-280 beads were added to lysates and incubated overnight rotating at 4°C.

High stringency washes and on-bead digestion for ChromID:

After incubation of nuclear lysates with beads rotating at 4°C overnight, beads were separated from the unbound fraction on a magnetic rack and washed twice with 2% SDS in TE (+ 1mM DTT, 1x PIC) for 10 minutes rotating overhead at room temperature (RT), once with high salt buffer (HSB; 50mM HEPES pH 7.5, 1mM EDTA, 1% Triton X-100, 0.1% deoxycholate, 0.1% SDS, 500mM NaCl, 1mM DTT, and 1x PIC) for 10 minutes at RT, once with DOC buffer (250mM LiCl, 10mM Tris pH 8.0, 0.5% NP40, 0.5% deoxycholate, 1mM EDTA, 1mM DTT, and 1x PIC) for 10 minutes at 4°C, and twice with TE buffer (+ 1mM DTT, 1x PIC) for 10 minutes at 4°C. After the washes, beads were isolated from the last TE wash on a magnetic rack and the proteins were digested with 0.5 µg trypsin (Promega; V5111) in 40ul digestion buffer (1M Urea in 50mM Tris pH 8.0, 1mM Tris-(2-carboxyethyl)-phosphin (TCEP)) directly on beads, overnight at 26°C and shaking at 600 rpm. Next day, the digested protein-peptide mix was isolated from beads and reduced with 2mM TCEP for 45 minutes at RT, and then alkylated with 10mM Chloroacetamide (ClAA) for 30min at RT in the dark. The digestion was stopped the next day by acidifying the peptides with Trifluoracetic acid (TFA) to a final concentration of 0.5%, and the Acetonitrile (ACN) concentration was adjusted to 3% prior loading on C18 StageTips.

C18 StageTips clean-up:

Obtained petides were cleaned-up using in-house produced (Functional Genomics Center Zurich, FGCZ) C18-StageTips. First, these were humidified with 100% methanol (MeOH), cleaned twice with 60% ACN; 0.1% TFA, and conditioned twice with 3% ACN; 0.1% TFA. Then, peptides were loaded onto the StageTips, and the collected flow-through was loaded again. Afterwards, the peptides were desalted twice with 3% ACN; 0.1% TFA, and finally eluted twice with 60% ACN; 0.1% TFA. Desalted peptides were shock frozen in liquid nitrogen

(N2), completely dried in a speed vacuum centrifuge, and subsequently resolved in 3% ACN; 0.1% formic acid (FA), containing internal retention time (iRTs, Biognosys) standard peptides.

Detection of biotinylated proteins by data-dependent acquisition (DDA) mass spectrometry:
We used an Easy-nLC 1000 HPLC system operating in trap / elute mode (trap column: Acclaim PepMap 100 C18, 3um, 100A, 0.075x20mm; separation column: EASY-Spray C18, C18, 2um, 100A, 0.075x500mm, Temp: 50°C) coupled to an Orbitrap Fusion mass spectrometer (Thermo Scientific). Trap and separation column were equilibrated with 12 ul and 6 ul solvent A (0.1 % FA in water), respectively. 2 µl of the resuspended sample solution was injected onto the trap column at constant pressure (500 bar) and peptides were eluted with a flow rate of 0.3 µl/min using the following gradient: 2 % - 25 % B (0.1 % FA in ACN) in 50 min, 25 % - 32 % B in 10 min an 32 % - 97 % B in 10 min. After 10 min of washing by 97 % B. High accuracy mass spectra were acquired with an Orbitrap Fusion mass spectrometer (Thermo Scientific) using the following parameter: scan range of 300-1500 m/z, AGC-target of 4e5, resolution of 120'000 (at m/z 200), and a maximum injection time of 50 ms . Data-dependent MS/MS spectra were recorded in top speed mode in the linear ion trap using quadrupole isolation (1.6 m/z window), AGC target of 1e4, 300 ms maximum injection time, HCD-fragmentation with 30 % collision energy, a maximum cycle time of 3 sec, and all available parallelizable time was enabled. Mono isotopic precursor signals were selected for MS/MS with charge states between 2 and 7 and a minimum signal intensity of 5e3. Dynamic exclusion was set to 25 sec and an exclusion window of 10 ppm. After data collection, the peak lists were generated using automated rule-based converter control [73] and Proteome Discoverer 2.1 (Thermo Scientific).

Protein identification and label-free protein quantification of DDA data:
Protein identification and label-free quantification was done with MaxQuant (version 1.5.3.30) using the Andromeda search engine and label-free quantification (LFQ) [74]. The mouse reference proteome (UniProtKB/Swiss-Prot and UniProtKB/TrEMBL) version 2018_12

927 combined with manually annotated contaminant proteins was searched with protein and
928 peptide FDR values set to 0.01. All MaxQuant parameters can be found in the uploaded
929 parameterfile: rpx40_mqpar.xml (deposited in the PRIDE repository). Perseus (versions
930 1.6.1.1) was used for statistical analysis [75]. Results were filtered to remove reverse hits and
931 proteins only identified by site. Further, only proteins found in at least 3 replicates were kept.
932 Missing values were imputed from a 1.8 standard deviations left-shifted Gaussian distribution
933 with a width of 0.3 (relative to the standard deviation of measured values). Potential interactors
934 were determined using a t-test and visualised by a volcano plot. Significance lines were
935 determined by a permutation-based method [76] with an FDR value of 0.01 and S0 values (curve
936 bend) of 0.1 to 1 (details shown in volcano plots). Obtained results were exported and further
937 visualised using the statistical computer language R (version 3.5.2).
938
939 Estimation of protein abundance by data-independent acquisition (DIA)
940 Cells were grown to about 90% confluency on 10 cm dishes (approximately 15 x 10^6 cells),
941 harvested by trypsinisation, and pelleted by centrifugation at 1000 rpm for 5 min. Cell nuclei
942 were extracted following the nuclear extraction procedure (described above) until digestion of
943 genomic DNA. Nuclei were then pelleted by centrifugation at 2000 g for 10 minutes and
944 resuspended in 30 µl lysis buffer (4% (w/v) SDS, 100 mM Tris/HCL pH 8.2). Lysate was
945 incubated at 95 °C for 5 min under 1000 rpm shaking, followed by centrifugation at 16000 g
946 for 10 minutes at RT. Supernatant was processed immediately using FASP (Wisniewski et al,
947 2009) using 50 µg of total protein as measured by Qubit Protein Assay Kit (Q33211,
948 ThermoFisher). Tryptic peptides were cleaned-up using in-house produced C18-StageTips.
949 Peptides were resuspended in 3% acetonitrile, 0.1% formic acid in water including iRT
950 peptides (Biognosys, Schlieren, Switzerland).
951 All Data-independent acquisition (DIA) was performed on an Orbitrap Fusion Lumos mass
952 spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled online to a Acquity UPLC M-
953 class (Waters, Milford, MA) using a PicoView 565 nanospray source (New Objective). Peptide
954 mixtures were separated in a single-pump trap/elute mode, using a trapping (nanoEase

Symmetry C18, 5 µm, 180 µm × 20 mm) and an analytical column (nanoEase HSS T3 C18, 100 Å, 1.8 µm, 75 µm x 250 mm). Solvent A was water, 0.1% formic acid and solvent B was acetonitrile, 0.1% formic acid. 0.5 µg peptides/sample were loaded with a constant flow of 0.5% solvent B, at 15 µL/min onto the trapping column. Trapping time was 0.5 min. Peptides were eluted via the analytical column with a constant flow of 300 nL/min. During the elution step, the percentage of solvent B increased in a nonlinear fashion from 8% to 22% in 82 min and 22% to 32% in 8 min.

In brief, following MS settings were applied: MS1 scan at 120'000 Orbitrap resolution with an AGC target of 1 × 10^6 and max. injection time of 118 ms in the mass range of 350 to 1205 m/z, followed by 50 DIA scans covering a precursor mass range of 400 to 1000 m/z with isolation window widths of 12 m/z. The scan resolution in the Orbitrap was set to 15'000 with an AGC target of 1 × 10^6 and max. injection time of 25 ms. The HCD collision energy was set to 30%.

Data Analysis. In brief, raw files were analysed in Spectronaut Pulsar (13.9.191106.43655, Biognosys) by library-free DirectDIA. The basic principles of DirectDIA analysis have been previously described by Tsou et al. (DOI: 10.1038/nmeth.3255). The searches were done against the mouse reference proteome (UniProtKB/Swiss-Prot) and the GFP reference. Search results were filtered at 1% FDR on precursor and protein group level. Only the top 3 peptides were used for label-free protein intensity calculation. The protein group report of significant proteins was further used for plotting in R.

Functional gene set enrichment and network visualisation:

All proteins identified previously were mapped to human STRING identifiers via the gene names and sequence similarity. Functional gene set enrichment was performed using the "Proteins with Values/Ranks" functionality in STRINGv11 [61] for each chromatin reader. The log2 fold changes over background were used. From all terms enriched in any of the chromatin readers, nine Gene Ontology Cellular Component terms (The Gene Ontology Consortium, 2019) which were significantly enriched in at least one of the readers were selected.

Cytoscape (version 3.7.1) was used to layout the 79 proteins that were identified in ChromID experiments and are members of at least one of the selected GO terms. Visualisation was based on GO term membership only. Each protein was represented by a pie chart which signifies in which reader the protein was significantly detected after LS-MS/MS. STRING interaction confidences were added as links between proteins, with a cutoff set at confidence 0.4. For foreground protein network visualisation, all proteins with a positive log2 fold change in any of the chromatin readers compared to nBASU were considered as foreground. Their protein-protein interaction network was retrieved from STRINGv11 [61] with an interaction confidence threshold of 0.7. The network was imported into Cytoscape (version 3.7.1) and visualized using the "Prefuse Force Directed OpenCL Layout".

Data access:

All sequencing datasets produced in this study have been deposited to the NCBI Gene Expression Omnibus under the accession: GSE128907. Additional sequencing datasets from other publications are listed in the Supplementary Table 3. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD014483.

Figure Legends

Figure 1 - Nuclear and genomic localisation indicates correct eCR interactions with the genome

a) Top: chromatin reader domains used in this study and their specificities towards chromatin marks. Bottom: schematics of constructs utilised to generate mouse embryonic stem cells expressing engineered chromatin readers (eCRs). eCRs are composed of single or dual chromatin reader domains fused in frame to eGFP and a biotin acceptor site (see also Supplementary Fig. 1a). **b)** Recombinase-Mediated Cassette Exchange (RMCE) followed by double selection (Ganciclovir and Puromycin) was applied to generate stable integration of the expression construct at a defined site in the mouse genome. Stably expressed eCRs are *in vivo* biotinylated by a bacterial BirA ligase. **c)** Live imaging shows nuclear localisation of single and dual eCRs in mouse ES cells. Nuclear eGFP serves as control (See also Supplementary Fig. 2). Size bars = 5 µm. **d-f)** Genome browser examples showing correct localisation of eCRs according to chromatin modifications detected by antibody-ChIP-seq and DNA methylation by MeDIP-seq. For histone-PTM readers, only eCRs with two reader domains are shown (See supplementary Fig. 3 for single-domain eCRs). Shown is the library-normalised read density at 100bp intervals. Gene models and the position of CpG islands and repetitive elements are indicated. **g)** Pearson correlation score obtained from comparisons of eCRs with chromatin modifications at selected genomic intervals positive for the interrogated chromatin modifications.

Figure 2 - Functional analysis indicates dependency on reader domains and modifications for correct eCR localisation.

a-b) Genome browser examples for loss of binding of mutant-eCRs to sites enriched by wild type eCRs. Shown is the library-normalised read density at 100bp intervals. Gene models and the position of CpG islands and repetitive elements are indicated. **c)** Genome browser

example for loss of CBX7-2xChromo-eCR binding to the genome in absence of H3K27me3 in *Eed*-KO cells. **d)** Box plots showing loss of binding of mutant eCRs to sites bound by their wild-type versions. In addition, removal of binding substrates such as DNA methylation (*Dnmt*-TKO) or H3K27me3 (*Eed*-KO) results in loss of binding of wild-type eCRs to the corresponding sites. Shown are log2-FC enrichments at peak regions identified for the wild-type eCRs in wild type cells. Boxes denote the IQR and whiskers 1.5 IQR. **e)** Heatmap showing the histone-PTM log2-FC enrichment scores obtained from CBX1-2xChromo and TAF3-2xPhd eCR ChIP. Histone-PTMs are clustered based on enrichment scores. Colour code below the heatmap indicates histone isoforms and modification types detected in this assay. **f-g)** Bar plots showing enrichment/depletion for selected histone-PTMs, and the effect of combinatorial serine-10 phosphorylation on CBX1-2xChromo binding. Error bars indicate the standard deviation from two independent replicate measurements.

Figure 3 - ChromID identifies proteins associated with H3K9me3 and DNA methylation.

a) Schematic describing ChromID using engineered chromatin readers fused to promiscuous biotin ligases (BioL). **b)** Volcano plot showing ChromID results obtained using the CBX1-eCR-BASU targeting H3K9me3 over a reader-free nuclear BASU (nBASU) control. Statistically-enriched proteins are indicated (FDR-corrected t-test, FDR = 0.01, n = 4 independent replicates). Peptides used to identify CBX1 match the Chromodomain used in the eCR. **c)** Bar plots representing the top 10 cellular component GO terms enriched by CBX1-eCR-BASU. The combined score is calculated by multiplying the ln(p-value) from Fisher's exact test and the z-score. **d)** Heatmap representation of significantly enriched proteins captured with the 5mC-reader (MBD1-eCR-BASU) and compared to results obtained using a mutant 5mC-reader (R22A-eCR-BASU). Shown are average LFQ intensities (log2-FC) from four independent measurements. Peptides used to identify MBD1 match the MBD domain used in the eCR. **e)** Same as in c, but for proteins enriched by MBD1-eCR-BASU.

Figure 4 - Generation and validation of eCRs reading bivalent H3K4me3 and H3K27me3 marks.

a) Genome browser example showing context-dependent localisation of dual-reader eCRs to bivalent sites decorated by H3K4me3 and/or H3K27me3. Binding is preferentially directed to bivalent sites, while regions modified by H3K4me3- or H3K27me3-only show less recruitment. Gene models and the position of CpG islands and repetitive elements are indicated. **b)** Scatter plots indicating the distribution (highlighted data points) and enrichment (colour) for the tested eCRs along the mouse genome based on H3K27me3 and/or H3K4me3 marks. Shown is the enrichment of H3K27me3 and H3K4me3 at 1kb windows covering the entire genome (grey). Coloured data points indicate the top 1% genomic windows enriched by the indicated eCR. eCRs specific for one modification separate towards their respective substrates, while the dual reader localises predominantly to the bivalent-modified sites (See also Supplementary Fig. 11d). **c)** Average density profiles around H3K4me3- (red) and H3K27me3-monovalent (blue) or bivalent-peaks (green). Data indicates increased preference of the dual-reader eCR for the bivalent peaks while binding at H3K4me3- and H3K27me3-only peaks is strongly reduced (See also Supplementary Fig. 11e).

Figure 5 - ChromID identifies the proteome associated with key chromatin marks in mouse ES cells.

a) Bar plots representing the top 10 cellular component GO terms enriched at H3K4me3-, H3K27me3- and bivalently-modified chromatin. **b)** Genome browser example for FLASH/CASP8AP2 co-localising at transcribed histone genes marked with H3K4me3. **c)** Heatmap representation of proteins significantly-enriched in either of the ChromID experiments specific for H3K9me3-, H3K4me3-, H3K27me3- or bivalently-modified chromatin in mouse ES cells (FDR-corrected t-test, FDR = 0.01, n = 4 independent replicates). The LFQ intensities (log2-FC) over nBASU are shown. **d)** Network analysis based on proteins belonging to major cellular component GOs terms identified in all ChromID experiments. Individual proteins are shown as nodes, edges indicate interactions retrieved from the STRING database

(interaction score > 0.9). Proteins detected in ChromID experiments but not called significant by the t-test are shown as grey nodes. Significantly enriched proteins are coloured according to the reader they have been identified. **e)** Heatmap representation of identified factors classified based on their functionality and clustered according to the computed LFQ intensities (log2-FC/nBASU). Proteins were selected based on min 0.5 log2-FC in at least one ChromID experiment.

Supplementary Figures

Supplementary Figure 1

a) Constructs used to integrate and express eCRs in mouse stem cells. eCR variants are available using either one or two readers domains in tandem. Triangles: LoxP sites; CAG: CAGGS promoter; bio: biotin acceptor site; NLS: nuclear localisation signal; eGFP: enhanced green fluorescence protein; IRES: internal ribosomal entry site; PAC: Puromycin N-Acetyl-transferase; 2A: 2A self-cleavable peptide. The 2A self-cleavable peptide enables to generate eCRs lacking GFP, which we used to test the influence of GFP on genome-wide eCR binding in Supplementary Fig. 2c. **b)** FACS profiles showing stable and homogenous expression of various eCRs in mouse ES cell lines. **c)** Barplots displaying protein abundances in eCR cell lines measured by direct-DIA. Top 3 precursors per peptide were used for quantity calculation of reference proteins (grey) and eCR-GFP fusion protein (green). Proteins were ordered from high to low protein abundance. **d)** Western blot analysis comparing protein levels of endogenous CBX1 and CBX1-eCR using an antibody specific for the CBX1 Chromodomain in nuclear extracts obtained from ES cells expressing either the eGFP-only or the CBX1-2xChromo-eCR constructs. LaminB1 is used as a loading control. **e-f)** FACS measurements of ES cell-derived neuronal progenitor cells after embryoid body dissociation shows identical percentage and distribution of cells expressing the neuronal surface marker CD24. ES cells were used as negative controls. e) indicates the percentage of positive cells and f) the CD24 intensity in the cell population. **g)** Microscopy images of *in vitro* derived neurons shows successful differentiation of neuronal cells derived from control ES cells and ES cells expressing eCRs.

Supplementary Figure 2

a) Live-cell imaging showing nuclear localization of single and dual eCRs. Nuclear eGFP serves as control and shows exclusion from mitotic chromosomes. Size bar = 5 µm. **b)** Fixed-

cell imaging comparing H3K9me3 localisation and CBX1-2xChromo eCR localisation by antibody staining. **c)** Comparison between fixed-cell and live-cell imaging using the CBX1-2xChromo eCR highlights the suitability of eCRs to study localisation on M-phase chromosomes. **d)** Live-cell imaging examples of stably-expressed dual eCRs in mouse ES cells showing localization to interphase chromosomes. Nuclear eGFP serves as control and shows exclusion from mitotic chromosomes. Size bar = 5 µm. **e)** Typical filmstrips of time-lapse fluorescence imaging of proliferating cells in the GFP control cell line (Top, from movie S1) and the H3K27me3 reader (CBX7-eGFP-eCR) expressing cell line (Bottom, from movie S2). Cells undergoing division are circled in yellow and were imaged at 5-min intervals. Size bar = 5 µm. **f)** Western blot detection of His-tagged eCRs enriched using reconstituted nucleosomes modified by either H3K4me3 or H3K27me3. Experiments were performed with three different amounts of the indicated eCRs expressed from *E.coli*. Left panel shows pulldowns performed using H3K4me3-modified mononucleosomes against TAF3-1xPhd or TAF3-2xPhd eCRs, and right panel using H3K27me3 against CBX7-1xChromo or 2xChromo.

Supplementary Figure 3

a-b) Genome browser examples for 1 x domain-eCR and 2 x domain-eCR localisation according to chromatin modifications and full-length proteins containing the reader domains (where available). Histone modifications were detected based on antibody-ChIP-seq. Shown is the library-normalised read density at 100bp intervals. Gene models and the position of CpG islands and repetitive elements are indicated. **c)** Cross-correlation matrix showing the genome-wide association of eCRs with histone modifications. Pearson's correlation score was calculated based on read counts at 1 kb intervals covering the entire genome. Regions with low mappability scores were removed from this analysis. **d)** Scatterplots showing binding of the eCRs to all analysed chromatin marks. Data points represent log2-transformed read counts at 1 kb intervals along the mouse genome. Pearson's correlation coefficient is indicated.

Supplementary Figure 4

a) Top: Scatterplots comparing binding of eCRs to binding data obtained from the full-length proteins containing the respective reader domains. Bottom: Scatterplots comparing binding of the full-length proteins to the respective chromatin modifications. Data points represent log2-transformed read counts at 1 kb intervals along the mouse genome. Pearson's correlation coefficient is indicated. **b)** Heat maps of histone modifications and eCR binding to peaks identified based on antibody ChIP-seq data for H3K4me3, H3K27me3 and H3K9me3. These results indicate a good agreement in binding of eCRs to the corresponding marks. Note that H3K4me3 and H3K27me3 have a substantial overlap due to bivalently modified regions. **c)** Box plots showing similar enrichments of eCRs and histone modifications at peaks identified using H3K4me3, H3K27me3 and H3K9me3 data sets (same as in b). Enrichment is calculated as log2-fold change of ChIP-seq reads over corresponding input reads. Boxes denote the inter-quartile range (IQR) and whiskers 1.5 ICR. **d)** Bar plots showing the percentage of histone modification peaks (red:H3K4me3, green:H3K9me3, blue:H3K27me3) covered by peaks identified in the eCR ChIP-seq data. Increased coverage of H3K4me3 peaks by the H3K27me3 readers (CBX7-eCR, dPC-eCR) is due to the strong overlap between H3K27me3 and H3K4me3 peaks at bivalent sites in mouse ES cells. Coverage was calculated based on base pairs overlapping the indicated peak sets.

Supplementary Figure 5

a-b) Box plots showing enrichment of chromatin marks at peaks identified using individual eCR data sets. Histone modification enrichment is calculated as log2-fold change of ChIP-seq reads over input reads. DNA methylation is calculated as percentage of methylated cytosines in the peak region using whole genome bisulphite sequencing (WGBS) data. Boxes denote the inter-quartile range (IQR) and whiskers 1.5 ICR. **c)** ChromHMM-based seven-state segmentation of the mouse ES cell genome based on four histone marks. Segmentation was obtained from ENCODE/UCSC and table indicates the model parameters used. **d)** Box plots showing enrichment of chromatin marks and eCRs at the seven states obtained from

ChromHMM. Enrichments are calculated as log2-fold change of ChIP-seq reads over input reads. Boxes denote the inter-quartile range (IQR) and whiskers 1.5 ICR. **e)** Stacked bar plots showing the eCR and histone modification peak-coverage distribution among various genomic features. Coverage was calculated based on base pairs overlapping between the peaks and the indicated features. “Genome-wide” the coverage proportions of the used features in the mouse genome. **f)** Scatterplots showing increased binding specificity of eCRs containing two reader domains to their binding substrates. Data points represent log2-fold changes calculated based on read counts at 1 kb intervals of 2 x domain eCRs over 1 x domain eCRs (y-axis) versus log2-fold enrichment of ChIP over input for histone marks (x-axis). Pearson’s correlation is indicated.

Supplementary Figure 6

a) FACS profiles comparing expression levels and expression homogeneity in the cell population for cell lines expressing wild type and mutant eCR. **b)** Immunofluorescence indicates localisation of wild type and mutant eCRs in wild type ES cells and wild type eCRs in mutant cell lines (*Eed*-KO and *Dnmt*-TKO). Size bar = 5 µm. **c)** Genome browser example for loss of binding of mutant-MBD-eCRs to sites enriched by wild type eCRs. Shown is the library-normalised read density at 100bp intervals. **d)** HPLC-MS measurement of methylcytosine and cytosine indicates absence of methylation in the *Dnmt*-TKO clones used here. Error bars denote standard deviation from three independent replicate measurements. **e)** Western blot analysis validates loss of H3K27me3 in the *Eed*-KO cell lines. Histone H1 and H3 serve as loading controls. **f)** Scatterplots showing loss of binding to chromatin substrates for wild CBX7-Chromo- or MBD-based eCRs in cell lines lacking H3K27me3 or DNA methylation, respectively. Shown are log2-fold changes between eCRs mapped in wild type and mutant ES cells (y-axis) versus the respective modification in wild type cells (x-axis) either as log2-fold enrichment of histone modifications over input chromatin, or local concentration of methyl-CpGs (log2 / 100bp based on WGBS data) for MBD-eCRs. Data points are calculated based on 1kb intervals.

Supplementary Figure 7

a) Schematics indicating the histone-PTM quantification procedure, involving eCR-dependent chromatin enrichment by ChIP, followed by PAGE purification of histones, and MS-based detection of histone-PTMs (See Methods). **b)** Brilliant Blue-stained PAGE gels indicate separation of histones obtained from eCR pulldowns (elute) and corresponding input chromatin after de-crosslinking. Red rectangle indicates the size-range range used to excise histone bands from the gel. c) Western blot detection of His-tagged TAF3-2xPhd eCRs enriched using reconstituted nucleosomes unmodified or modified by either H3K4me1 or H3K4me3. Anti-Histone H4 serves as a pull-down control. **d)** Scatter plots indicating the distribution (highlighted data points) and enrichment (colour) for the indicated eCRs along the mouse genome based on H3K9me3 and/or H4K20me3 marks. Shown is the enrichment of H3K9me3 and H4K20me3 at 1kb windows covering the entire genome (grey). Coloured data points indicate the top 1% genomic windows enriched by the indicated eCR. The CBX1-eCR localises predominantly to sites that contain H3K9me3, irrespective of presence of H4K20me3. **b)** Boxplots indicating CBX1-2xChromo and H3K9me3 (AbChIP) enrichments at genomic intervals containing H3K9me3 & H4K20me3 and at sites that contain H4K20me3 in absence of H3K9me3.

Supplementary Figure 8

a) Constructs used to integrate and express eCRs fused to promiscuous biotin ligases in mouse stem cells. Control constructs consist of nuclear reader-free promiscuous biotin ligases. Triangles: LoxP sites; CAG: CAGGS promoter; bio: biotin acceptor site; NLS: nuclear localisation signal; 13xGGGGS: flexible long linker [33]; BioL: promiscuous biotin ligase; 2A: 2A self-cleavable peptide; eGFP: enhanced green fluorescence protein; IRES: internal ribosomal entry site; PAC: Puromycin N-Acetyl-transferase. **b)** Streptavidin Western blot to test the labelling efficiency of three different promiscuous biotin ligases without a chromatin reader domain (BirA* R118G-mutant, BioID2 and BASU) in mouse stem cells for the indicated time

points. **c)** Heatmap representation of significantly-enriched proteins captured with the H3K9me3-reader fused to BASU (CBX1-eCR-BASU) or BioID2 (CBX1-eCR-BioID2). Shown are LFQ intensities (log2-FC). **d)** Volcano plots showing ChromID results obtained using CBX1-eCR-BioID2 (left) or CBX1-eCR-BASU (right) targeting H3K9me3 over a reader-free nuclear BioID2 (nBioID2) or reader-free nuclear BASU (nBASU) control (24h induction). Statistically enriched proteins are indicated (FDR-corrected t-test, n = 4 individual replicates). **e)** Heatmap representation of enriched proteins (top) or background proteins (bottom) captured with the H3K9me3-reader fused to BASU (CBX1-eCR-BASU) after three different biotin induction times (1, 6 and 24 hours). Shown are LFQ intensities (log2-FC).

Supplementary Figure 9

a) Endogenous tagging strategy for Znf280D, allowing for the insertion of a Flag-bio peptide at the C-terminal end of the protein using Cas9 and donor ssDNA. **b)** Sanger sequencing to confirm correct in-frame insertion of the Flag-bio peptide sequence. **c)** Genome browser examples showing localisation of ZNF280D at sites positive for H3K9me3 and bound by the CBX1-2xChromo eCR. **d)** Average density profiles and heat maps over all ZNF280D peaks indicating increased H3K9me3 localisation at ZNF280D-bound sites.

Supplementary Figure 10

a) Volcano plots showing ChromID results obtained using MBD1-eCR-BASU (left) or the binding mutant 5mC-reader (R22A-eCR-BASU) over nBASU control (12h induction). Statistically enriched proteins are indicated (FDR-corrected t-test, n = 4 individual replicates). **b)** Principal component analysis for all ChromID samples shown in f) indicates clustering based on identified proteins. **c)** Heatmap representation of enriched proteins captured with MBD1-eCR-BASU, nBASU and the binding mutant 5mC-reader (R22A-eCR-BASU) after 12h biotin induction. Shown are LFQ intensities (log2-FC).

Supplementary Figure 11

a) Schematic indicating the setup of the dual-reader eCR, specific for H3K27me3 and H3K4me3. **b)** FACS profiles comparing expression levels and expression homogeneity in cell lines expressing wild type and mutant dual-reader eCRs. **c)** Genome browser example showing context-dependent localisation of dual-reader eCRs to bivalent sites decorated by H3K4me3 and H3K27me3. **d)** Scatter plots indicating the distribution (highlighted data points) and enrichment (colour) for the tested eCRs along the mouse genome based on H3K27me3 and/or H3K4me3 marks. Same representation as in Fig. 4b. **e)** Box plots showing enrichment of the bivalent eCRs at genomic regions containing only H3K27me3 or H3K4me3 and regions bivalently modified by H3K4me3 and H3K27me3. Regions were identified based on antibody-ChIP-seq peaks, specific for H3K4me3 or H3K27me3. Enrichments are calculated as log2-fold change of ChIP-seq reads over input reads. Boxes denote the inter-quartile range (IQR) and whiskers 1.5 ICR. **f)** Cross-correlation matrix based on 1kb genomic intervals covering the entire mouse genome, showing correlations between H3K27me3, H3K4me3 and monovalent eCRs.

Supplementary Figure 12

a) Genome browser example showing loss of localisation to bivalent sites for dual-reader eCRs containing mutated Cromo- or Phd-domains. **b)** Heat map summarising the correlation of all tested eCRs with CpG density, H3K4me3 and H3K27me3 at 1 kb windows covering the entire genome. **c)** Average density profiles around genomic regions enriched for H3K4me3-only (red), H3K27me3-only (blue) or bivalently modified by both marks (green), showing loss of binding for bivalent eCRs containing mutations in either the H3K27me3- or H3K4me3-reader domains. **d)** Genome browser example showing loss of localisation to bivalent sites for dual-reader eCRs in ES cell lines lacking H3K27me3. **e)** Average density profiles showing loss of binding for bivalent eCRs in absence of H3K27me3. Same as in c).

Supplementary Figure 13

a) List of significantly-enriched proteins detected in all ChromID experiments (FDR-corrected t-test, FDR = 0.01, n = 4 individual replicates). Asterisks indicate proteins with chromatin reader domains specific for the respective mark. **b)** Heatmap representation of the significantly enriched proteins captured with each eCR-BASU targeting H3K4me3-, H3K27me3- and bivalently-modified chromatin in mouse ES cells (FDR-corrected t-test, FDR = 0.01, n = 4 individual replicates). The LFQ intensities (log2-FC) over nBASU are shown. **c)** Box plots showing enrichment of the H3K9me3 and H3K27me3-specific eCRs at genomic regions enriched by either CBX1-2xChromo (green) or CBX7-2xChromo (blue). This indicates lack of cross-reactivity between H3K27me3 and H3K9me3 readers. Enrichments are calculated as log2-fold change of ChIP-seq reads over input reads. Boxes denote the inter-quartile range (IQR) and whiskers 1.5 ICR. **d-e)** Volcano plots showing ChromID results obtained using either **d)** the CBX7-2xChromo or **e)** bivalent CBX7-Chromo+TAF3-Phd in ES cells lacking H3K27me3 (*Eed*-KO). Statistically enriched proteins are indicated (FDR-corrected t-test, n = 4 individual replicates).

Supplementary Figure 14

a) Genome browser example showing identical localisation between the same eCRs fused to either eGFP or BASU. Shown are biotin-ChIP-seq datasets using the TAF3-2xPhd and the CBX7-Chromo+TAF3-Phd constructs. **b)** Cross-correlation matrix based on 1kb genomic intervals covering the entire mouse genome, showing correlations between eGFP or BASU eCRs, together with H3K27me3, H3K4me3 and H3K9me3. c) Scatterplots somparing binding of the eCRs fused to eGFP and BASU. Data points represent log2-transformed read counts at 1 kb intervals along the mouse genome. Pearson's correlation coefficient is indicated.

Supplementary Figure 15

a) Principal component analysis based on detected protein intensities separates the nBASU controls and eCRs into distinct groups, with the H3K4me3 proteome and the H3K9me3 proteome showing the largest variation. The H3K9me3 and H3K27me3 datasets group closer

to each other, and the bivalent dataset forms a separate group between the H3K4me3 and H3K27me3, suggesting overlapping proteomes. **b)** Venn diagram showing the overlap of all significantly enriched proteins using eCRs specific for DNA methylation and H3K9me3.

References

1. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17,** 487–500 (2016).
2. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14,** 1025–1040 (2007).
3. Musselman, C. A., Lalonde, M.-E., Côté, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* **19,** 1218–1227 (2012).
4. Ruthenburg, A. J., Li, H., Patel, D. J. & David Allis, C. Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* **8,** 983–994 (2007).
5. Vermeulen, M. *et al.* Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. *Cell* **142,** 967–980 (2010).
6. Bartke, T. *et al.* Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation. *Cell* **143,** 470–484 (2010).
7. Nikolov, M. *et al.* Chromatin Affinity Purification and Quantitative Mass Spectrometry Defining the Interactome of Histone Modification Patterns. *Molecular & Cellular Proteomics* **10,** M110.005371–17 (2011).
8. Eberl, H. C., Spruijt, C. G., Kelstrup, C. D., Vermeulen, M. & Mann, M. A Map of General and Specialized Chromatin Readers in Mouse Tissues Generated by Label-free Interaction Proteomics. *Molecular Cell* **49,** 368–378 (2013).
9. Spruijt, C. G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152,** 1146–1159 (2013).
10. Mittler, G., Butter, F. & Mann, M. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome Research* **19,** 284–293 (2008).

11. DEjardin, J. & Kingston, R. E. Purification of proteins associated with specific genomic Loci. *Cell* **136,** 175–186 (2009).
12. Liu, X. *et al.* In Situ Capture of Chromatin Interactions by Biotinylated dCas9. 1–36 (2017). doi:10.1016/j.cell.2017.08.003
13. Myers, S. A. *et al.* Discovery of proteins associated with a predefined genomic locus via dCas9–APEX- mediated proximity labeling. *Nat Meth* 1–8 (2018). doi:10.1038/s41592-018-0007-1
14. Schmidtmann, E., Anton, T., Rombaut, P., Herzog, F. & Leonhardt, H. Determination of local chromatin composition by CasID. *Nucleus* **7,** 476–484 (2016).
15. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & Development* **17,** 1870–1881 (2003).
16. Bernstein, E. *et al.* Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell. Biol.* **26,** 2560–2569 (2006).
17. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410,** 120–124 (2001).
18. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410,** 116–120 (2001).
19. Vermeulen, M. *et al.* Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* **131,** 58–69 (2007).
20. Meehan, R. R., Lewis, J. D. & Bird, A. P. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Research* **20,** 5085–5092 (1992).
21. Nan, X., Meehan, R. R. & Bird, A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Research* **21,** 4886–4892 (1993).

22. Baubec, T., Ivanek, R., Lienert, F. & Schübeler, D. Methylation-dependent and - independent genomic targeting principles of the MBD protein family. *Cell* **153,** 480–492 (2013).

23. Peters, A. H. F. M. *et al.* Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nat Genet* **30,** 77–80 (2001).

24. Xu, J. *et al.* Super-Resolution Imaging of Higher-Order Chromatin Structures at Different Epigenomic States in Single Mammalian Cells. *CellReports* **24,** 873–882 (2018).

25. Hiragami-Hamada, K. *et al.* Dynamic and flexible H3K9me3 bridging via HP1β dimerization establishes a plastic state of condensed chromatin. *Nature Communications* **7,** 1–16 (2016).

26. Yap, K. L. *et al.* Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. *Molecular Cell* **38,** 662–674 (2010).

27. Jørgensen, H. F., Ben-Porath, I. & Bird, A. P. Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Mol. Cell. Biol.* **24,** 3387–3395 (2004).

28. Brown, K. *et al.* The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. *Human Molecular Genetics* **25,** 558–570 (2016).

29. Feller, C., Forné, I., Imhof, A. & Becker, P. B. Global and specific responses of the histone acetylome to systematic perturbation. *Molecular Cell* **57,** 559–571 (2015).

30. Schotta, G. *et al.* A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & Development* **18,** 1251–1262 (2004).

31. Fischle, W. *et al.* Regulation of HP1–chromatin binding by histone H3 methylation and phosphorylation. *Nat Cell Biol* **438,** 1116–1122 (2005).

32. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of Cell Biology* **196,** 801–810 (2012).

33. Kim, D. I. *et al.* An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell* **27,** 1188–1196 (2016).

34. Ramanathan, M. *et al.* RNA–protein interaction detection in living cells. *Nat Meth* **15,** 207–212 (2018).

35. Saksouk, N. *et al.* Erratum to Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation [Molecular Cell 56 (2014) 580-594]. *Molecular Cell* **57,** 202 (2015).

36. Daxinger, L. *et al.* An ENU mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. *Genome Biology* **14,** R96 (2013).

37. Tachibana, M. *et al.* G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes & Development* **16,** 1779–1791 (2002).

38. Peters, A. H. F. M. *et al.* Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *MOLCEL* **12,** 1577–1589 (2003).

39. Tchasovnikarova, I. A. *et al.* Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science* **348,** 1481–1485 (2015).

40. Goldberg, A. D. *et al.* Distinct Factors Control Histone Variant H3.3 Localizationat Specific Genomic Regions. *Cell* **140,** 678–691 (2010).

41. Agarwal, N. *et al.* MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. *Nucleic Acids Research* **35,** 5402–5408 (2007).

42. Arita, K. *et al.* Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 12950–12955 (2012).

43. Ueda, J., Tachibana, M., Ikura, T. & Shinkai, Y. Zinc finger protein Wiz links G9a/GLP histone methyltransferases to the co-repressor molecule CtBP. *Journal of Biological Chemistry* **281,** 20120–20128 (2006).

44. Nozawa, R.-S. *et al.* Human POGZ modulates dissociation of HP1α from mitotic chromosome arms through Aurora B activation. *Nature Publishing Group* **12,** 719–727 (2010).

45. Ambrosi, C., Manzo, M. & Baubec, T. Dynamics and Context-Dependent Roles of DNA Methylation. *Journal of Molecular Biology* **429,** 1459–1475 (2017).

46. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356,** eaaj2239–17 (2017).

47. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).

48. Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular Cell* **30,** 755–766 (2008).

49. Voigt, P. *et al.* Asymmetrically Modified Nucleosomes. *Cell* **151,** 181–193 (2012).

50. Yang, X.-C., Burch, B. D., Yan, Y., Marzluff, W. F. & Dominski, Z. FLASH, a Proapoptotic Protein Involved in Activation of Caspase-8, Is Essential for 3′ End Processing of Histone Pre-mRNAs. *MOLCEL* **36,** 267–278 (2009).

51. Boudreault, A. A. *et al.* Yeast enhancer of polycomb defines global Esa1-dependent acetylation of chromatin. *Genes & Development* **17,** 1415–1428 (2003).

52. Doyon, Y., Selleck, W., Lane, W. S., Tan, S. & Côté, J. Structural and Functional Conservation of the NuA4 Histone Acetyltransferase Complex from Yeast to Humans. *Mol. Cell. Biol.* **24,** 1884–1896 (2004).

53. Ravens, S., Yu, C., Ye, T., Stierle, M. & Tora, L. Tip60 complex binds to active Pol II promoters and a subset of enhancers and co-regulates the c-Myc network in mouse embryonic stem cells. *Epigenetics & Chromatin* 1–16 (2015). doi:10.1186/s13072-015-0039-z

54. Whetstine, J. R. *et al.* Reversal of Histone Lysine Trimethylation by the JMJD2 Family of Histone Demethylases. *Cell* **125,** 467–481 (2006).
55. Lu, C. *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483,** 474–478 (2012).
56. Das, P. P. *et al.* Distinct and Combinatorial Functions of Jmjd2b/Kdm4b and Jmjd2c/Kdm4c in Mouse Embryonic Stem Cell Identity. *MOLCEL* **53,** 32–48 (2014).
57. Horton, J. R. *et al.* Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat. Struct. Mol. Biol.* **17,** 38–43 (2010).
58. Fortschegger, K. *et al.* PHF8 Targets Histone Methylation and RNA Polymerase II To Activate Transcription. *Mol. Cell. Biol.* **30,** 3286–3298 (2010).
59. Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by Polycomb proteins. *Science Advances* **1,** e1500737–15 (2015).
60. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473,** 343–348 (2011).
61. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* **47,** D607–D613 (2019).
62. Li, H. *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442,** 91–95 (2006).
63. Kilic, S., Bachmann, A. L., Bryan, L. C. & Fierz, B. Multivalency governs HP1α association dynamics with the silent chromatin state. *Nature Communications* **6,** 1–11 (2015).
64. Brown, D. A. *et al.* The SET1 Complex Selects Actively Transcribed Target Genes via Multivalent Interaction with CpG Island Chromatin. *CellReports* **20,** 2313–2327 (2017).
65. Delachat, A. M. F. *et al.* Engineered Multivalent Sensors to Detect Coexisting Histone Modifications in Living Stem Cells. *Cell Chemical Biology* **25,** 51–56.e6 (2018).

66. Mauser, R., Kungulovski, G., Keup, C., Reinhardt, R. & Jeltsch, A. Application of dual reading domains as novel reagents in chromatin biology reveals a new H3K9me3 and H3K36me2/3 bivalent chromatin state. *Epigenetics & Chromatin* 1–19 (2017). doi:10.1186/s13072-017-0153-1

67. Tekel, S. J. *et al.* Tandem Histone-Binding Domains Enhance the Activity of a Synthetic Chromatin Effector. *ACS Synth. Biol.* **7,** 842–852 (2018).

68. Flemr, M. & Bühler, M. Single-Step Generation of Conditional Knockout Mouse Embryonic Stem Cells. *CellReports* 1–16 (2015). doi:10.1016/j.celrep.2015.06.051

69. Bibel, M., Richter, J., Lacroix, E. & Barde, Y.-A. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat Protoc* **2,** 1034–1043 (2007).

70. Derrien, T. *et al.* Fast Computation and Applications of Genome Mappability. *PLoS ONE* **7,** e30377–16 (2012).

71. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39,** 311–318 (2007).

72. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nature Publishing Group* **12,** 2478–2492 (2017).

73. Barkow-Oesterreicher, S., Türker, C. & Panse, C. FCC – An automated rule-based processing tool for life science data. *Source Code for Biology and Medicine 2013 8:1* **8,** 3 (2013).

74. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **26,** 1367–1372 (2008).

75. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Meth* **13,** 731–740 (2016).

76. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 5116–5121 (2001).

Villasenor et al. Figure 1

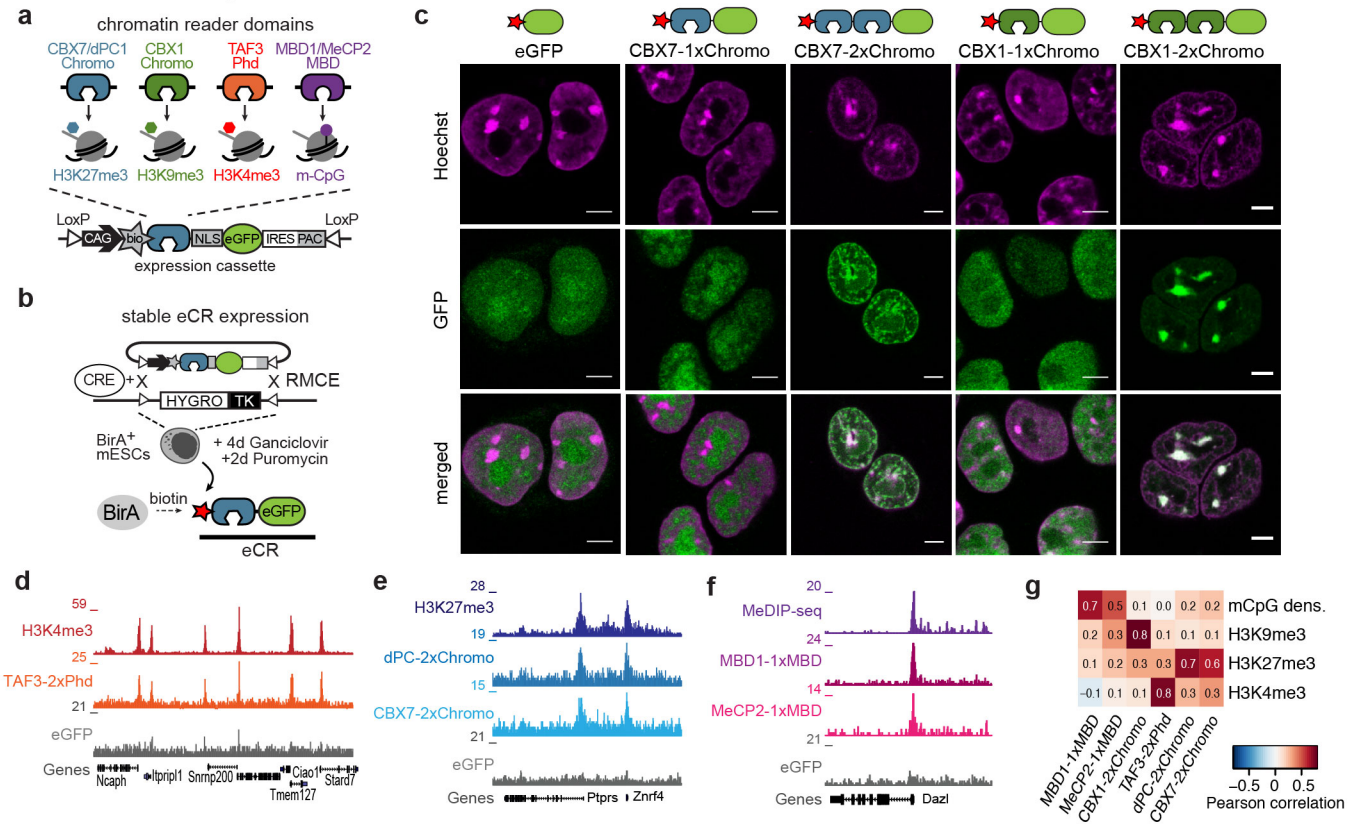


Figure 1 - Nuclear and genomic localisation indicates correct eCR interactions with the genome
a) Top: chromatin reader domains used in this study and their specificities towards chromatin marks. Bottom: schematics of constructs utilised to generate mouse embryonic stem cells expressing engineered chromatin readers (eCRs). eCRs are composed of single or dual chromatin reader domains fused in frame to eGFP and a biotin acceptor site (see also Supplementary Fig. 1a). b) Recombinase-Mediated Cassette Exchange (RMCE) followed by double selection (Ganciclovir and Puromycin) was applied to generate stable integration of the expression construct at a defined site in the mouse genome. Stably expressed eCRs are in vivo biotinylated by a bacterial BirA ligase. c) Live imaging shows nuclear localisation of single and dual eCRs in mouse ES cells. Nuclear eGFP serves as control (See also Supplementary Fig. 2). Size bars = 5 µm. d-f) Genome browser examples showing correct localisation of eCRs according to chromatin modifications detected by antibody-ChIP-seq and DNA methylation by MeDIP-seq. For histone-PTM readers, only eCRs with two reader domains are shown (See supplementary Fig. 3 for single-domain eCRs). Shown is the library-normalised read density at 100bp intervals. Gene models and the position of CpG islands and repetitive elements are indicated. g) Pearson correlation score obtained from comparisons of eCRs with chromatin modifications at selected genomic intervals positive for the interrogated chromatin modifications.

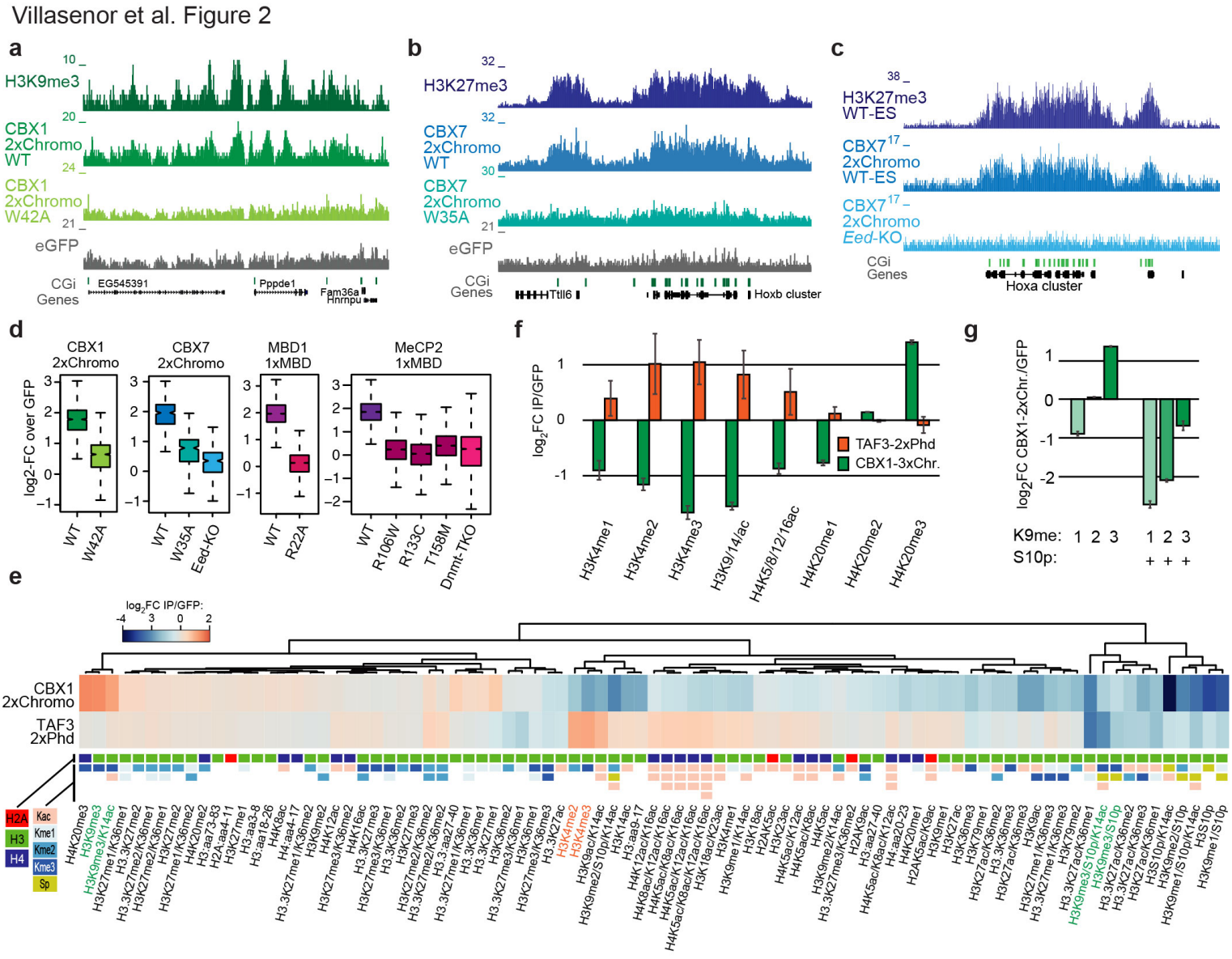


Figure 2 - Functional analysis indicates dependency on reader domains and modifications for correct eCR localisation.

a-b) Genome browser examples for loss of binding of mutant-eCRs to sites enriched by wild type eCRs. Shown is the library-normalised read density at 100bp intervals. Gene models and the position of CpG islands and repetitive elements are indicated. c) Genome browser example for loss of CBX7-2xChromo-eCR binding to the genome in absence of H3K27me3 in Eed-KO cells. d) Box plots showing loss of binding of mutant eCRs to sites bound by their wild-type versions. In addition, removal of binding substrates such as DNA methylation (Dnmt-TKO) or H3K27me3 (Eed-KO) results in loss of binding of wild-type eCRs to the corresponding sites. Shown are log2-FC enrichments at peak regions identified for the wild-type eCRs in wild type cells. Boxes denote the IQR and whiskers 1.5 IQR. e) Heatmap showing the histone-PTM log2-FC enrichment scores obtained from CBX1-2xChromo and TAF3-2xPhd eCR ChIP. Histone-PTMs are clustered based on enrichment scores. Colour code below the heatmap indicates histone isoforms and modification types detected in this assay. f-g) Bar plots showing enrichment/depletion for selected histone-PTMs, and the effect of combinatorial serine-10 phosphorylation on CBX1-2xChromo binding. Error bars indicate the standard deviation from two independent replicate measurements.

Villasenor et al., Figure 3

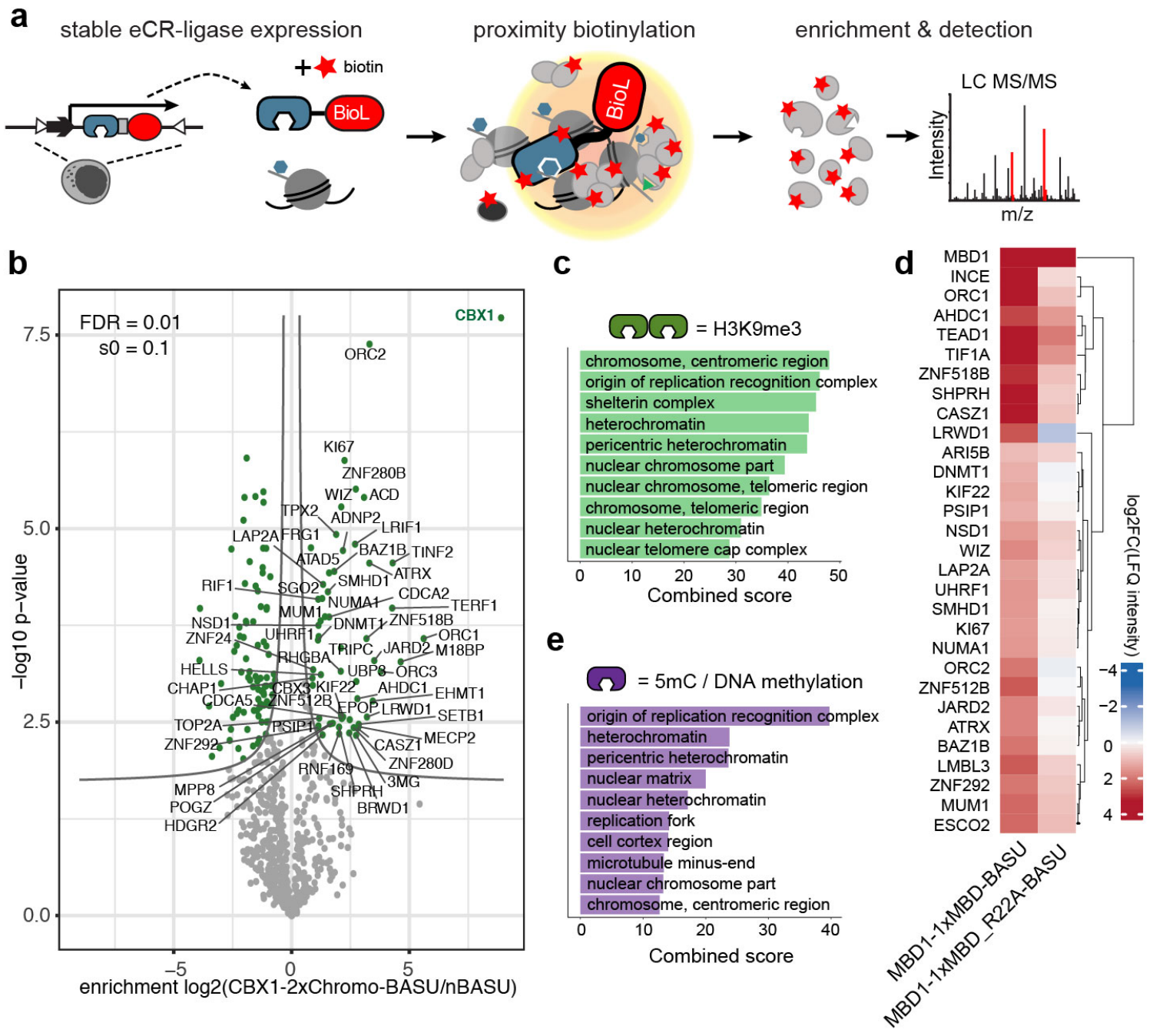


Figure 3 - ChromID identifies proteins associated with H3K9me3 and DNA methylation.
a) Schematic describing ChromID using engineered chromatin readers fused to promiscuous biotin ligases (BioL). b) Volcano plot showing ChromID results obtained using the CBX1-eCR -BASU targeting H3K9me3 over a reader-free nuclear BASU (nBASU) control. Statistically-enriched proteins are indicated (FDR-corrected t-test, FDR = 0.01, n = 4 independent replicates). Peptides used to identify CBX1 match the Chromodomain used in the eCR. c) Bar plots representing the top 10 cellular component GO terms enriched by CBX1-eCR-BASU. The combined score is calculated by multiplying the ln(p-value) from Fisher's exact test and the z-score. d) Heatmap representation of significantly enriched proteins captured with the 5mC-reader (MBD1-eCR-BASU) and compared to results obtained using a mutant 5mC-reader (R22A-eCR-BASU). Shown are average LFQ intensities (log2-FC) from four independent measurements. Peptides used to identify MBD1 match the MBD domain used in the eCR.
e) Same as in c, but for proteins enriched by MBD1-eCR-BASU.

Villasenor et al. Figure 4

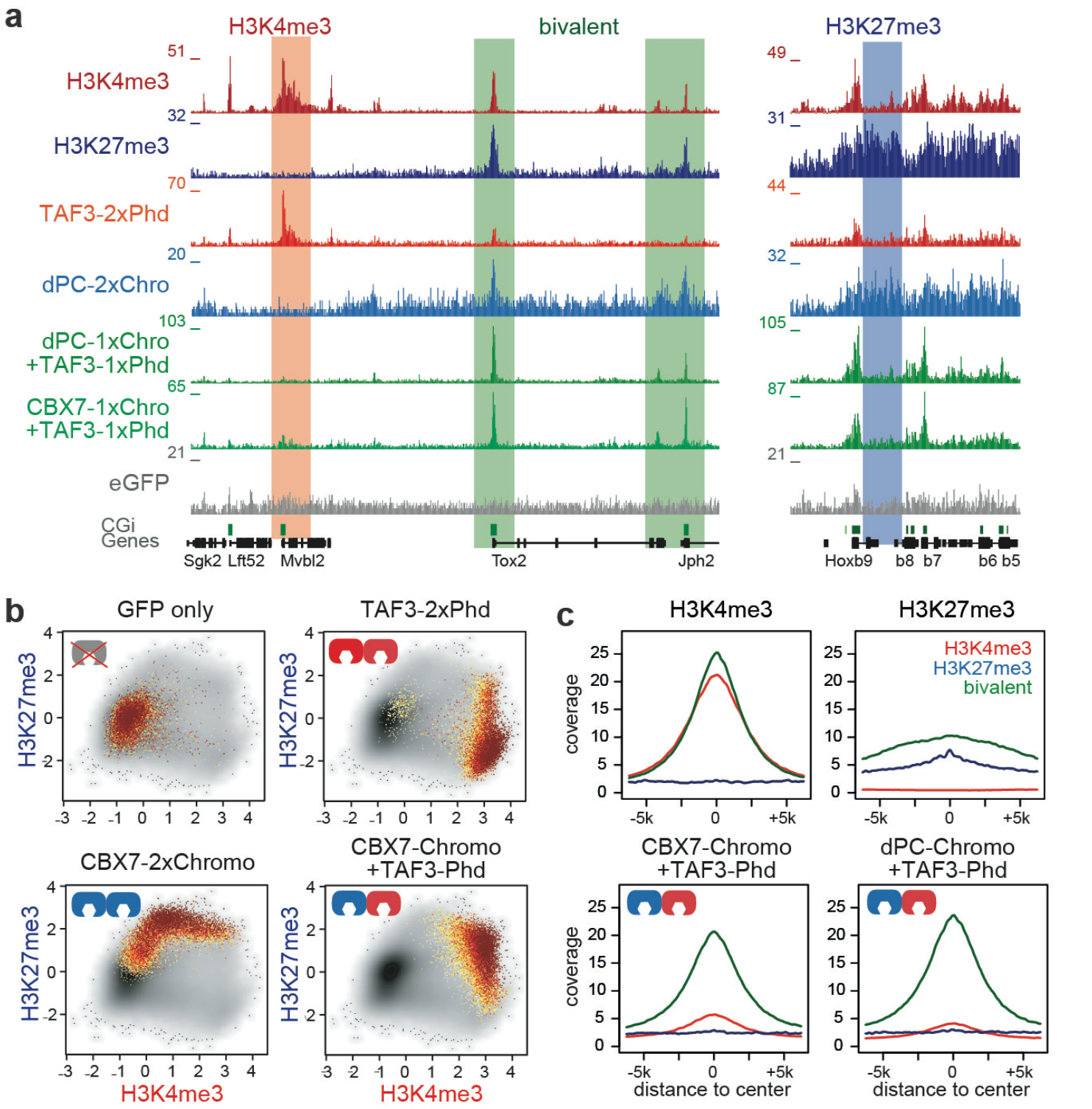


Figure 4 - Generation and validation of eCRs reading bivalent H3K4me3 and H3K27me3 marks. a) Genome browser example showing context-dependent localisation of dual-reader eCRs to bivalent sites decorated by H3K4me3 and/or H3K27me3. Binding is preferentially directed to bivalent sites, while regions modified by H3K4me3- or H3K27me3-only show less recruitment. Gene models and the position of CpG islands and repetitive elements are indicated. b) Scatter plots indicating the distribution (highlighted data points) and enrichment (colour) for the tested eCRs along the mouse genome based on H3K27me3 and/or H3K4me3 marks. Shown is the enrichment of H3K27me3 and H3K4me3 at 1kb windows covering the entire genome (grey). Coloured data points indicate the top 1% genomic windows enriched by the indicated eCR. eCRs specific for one modification separate towards their respective substrates, while the dual reader localises predominantly to the bivalent-modified sites (See also Supplementary Fig. 11d). c) Average density profiles around H3K4me3- (red) and H3K27me3-monovalent (blue) or bivalent-peaks (green). Data indicates increased preference of the dual-reader eCR for the bivalent peaks while binding at H3K4me3- and H3K27me3-only peaks is strongly reduced (See also Supplementary Fig. 11e).

Villasenor et al., Figure 5

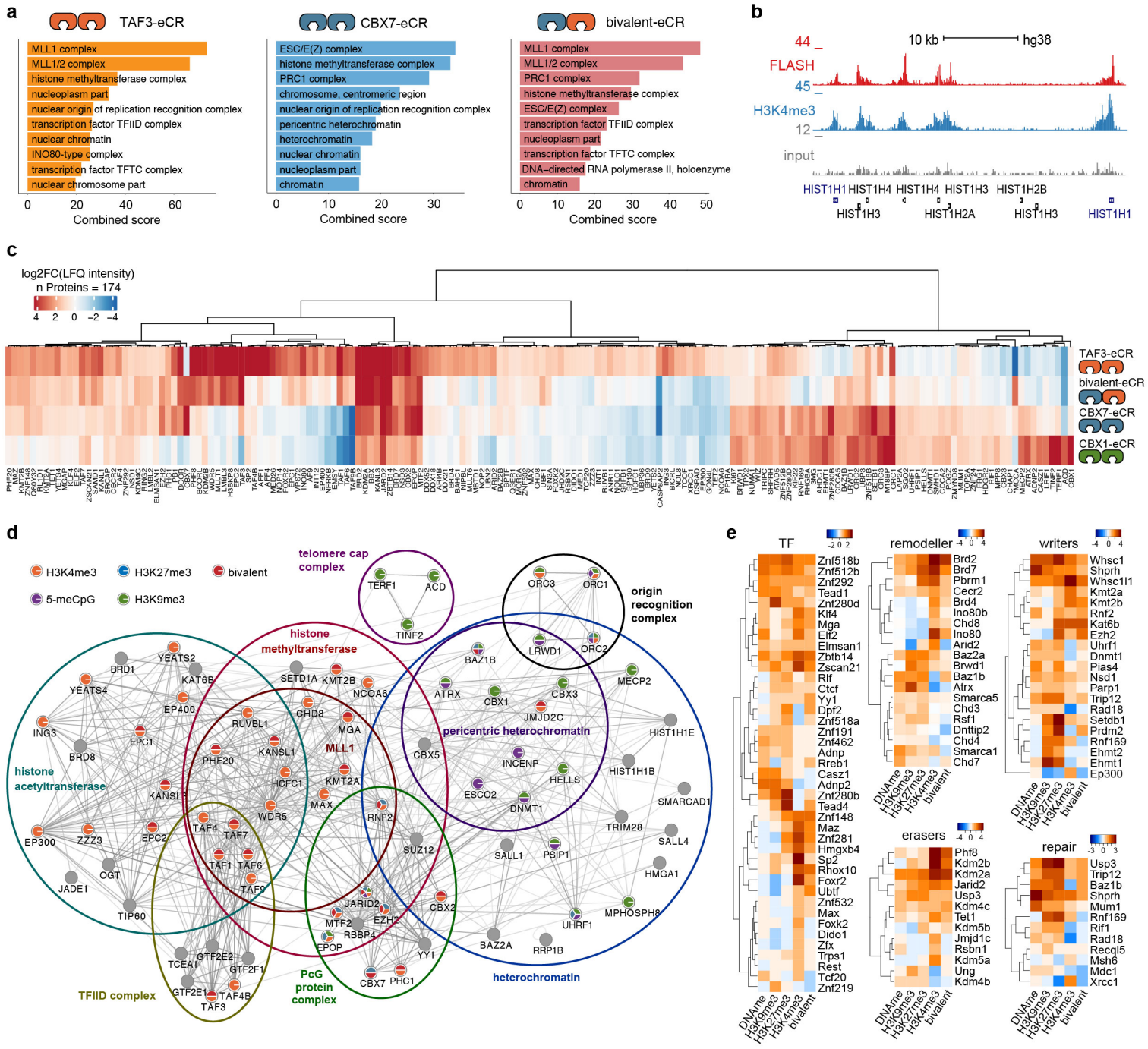
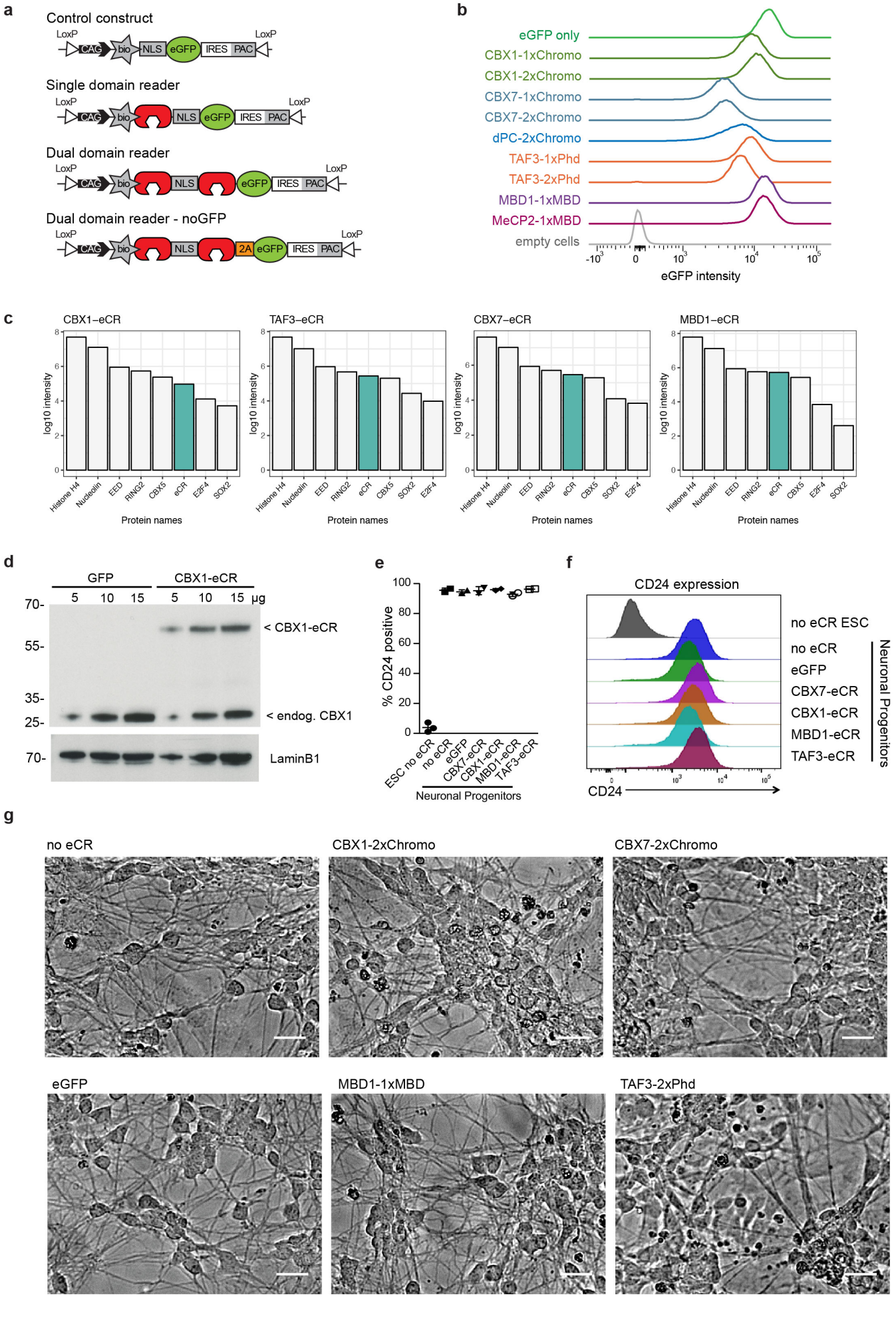
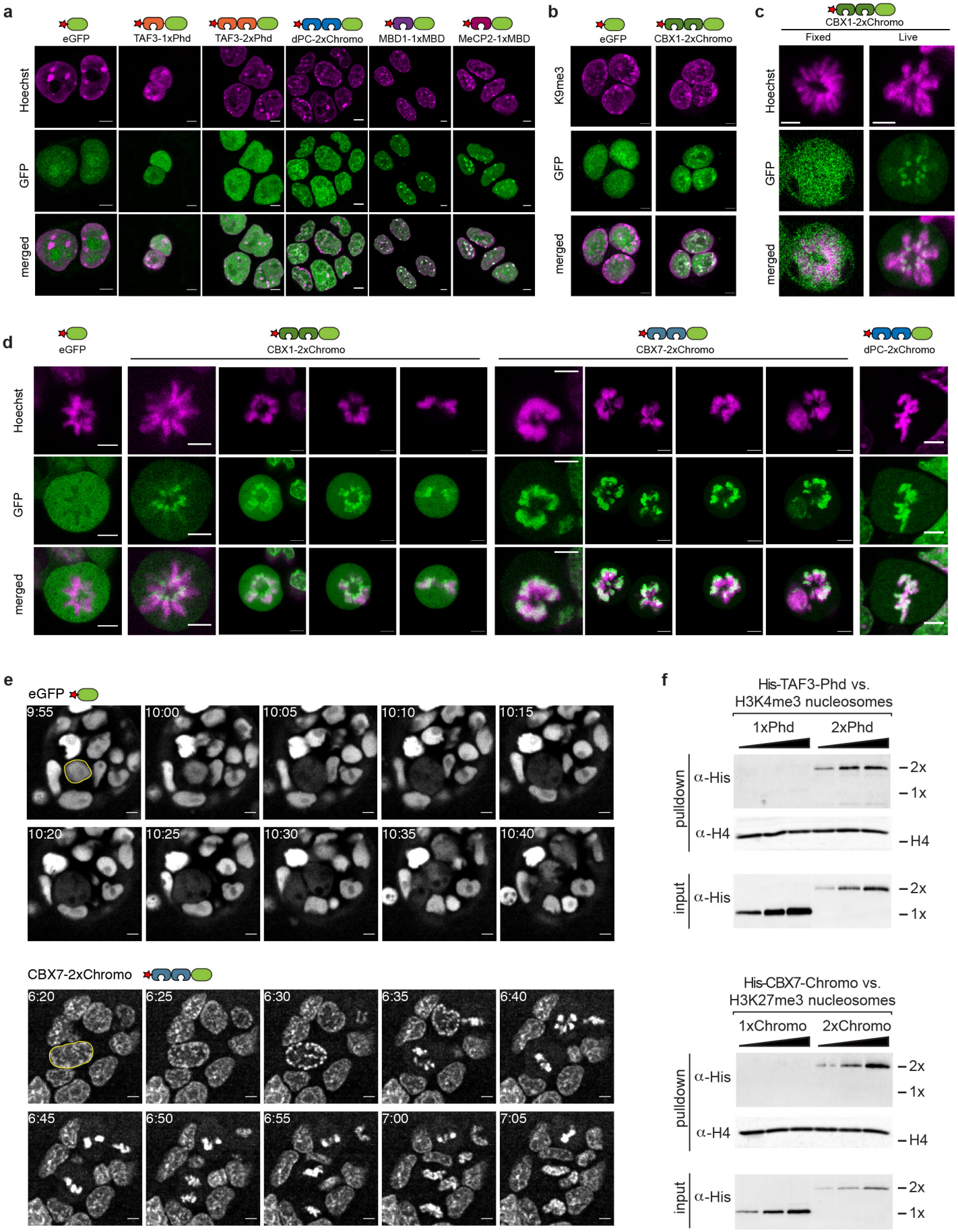


Figure 5 - ChromID identifies the proteome associated with key chromatin marks in mouse ES cells.
a) Bar plots representing the top 10 cellular component GO terms enriched at H3K4me3-, H3K27me3- and bivalently-modified chromatin. b) Genome browser example for FLASH/CASP8AP2 co-localising at transcribed histone genes marked with H3K4me3. c) Heatmap representation of proteins significantly-enriched in either of the ChromID experiments specific for H3K9me3-, H3K4me3-, H3K27me3- or bivalently-modified chromatin in mouse ES cells (FDR-corrected t-test, FDR = 0.01, n = 4 independent replicates). The LFQ intensities (log2-FC) over nBASU are shown. d) Network analysis based on proteins belonging to major cellular component GOs terms identified in all ChromID experiments. Individual proteins are shown as nodes, edges indicate interactions retrieved from the STRING database (interaction score > 0.9). Proteins detected in ChromID experiments but not called significant by the t-test are shown as grey nodes. Significantly enriched proteins are coloured according to the reader they have been identified.
e) Heatmap representation of identified factors classified based on their functionality and clustered according to the computed LFQ intensities (log2-FC/nBASU). Proteins were selected based on min 0.5 log2-FC in at least one ChromID experiment.

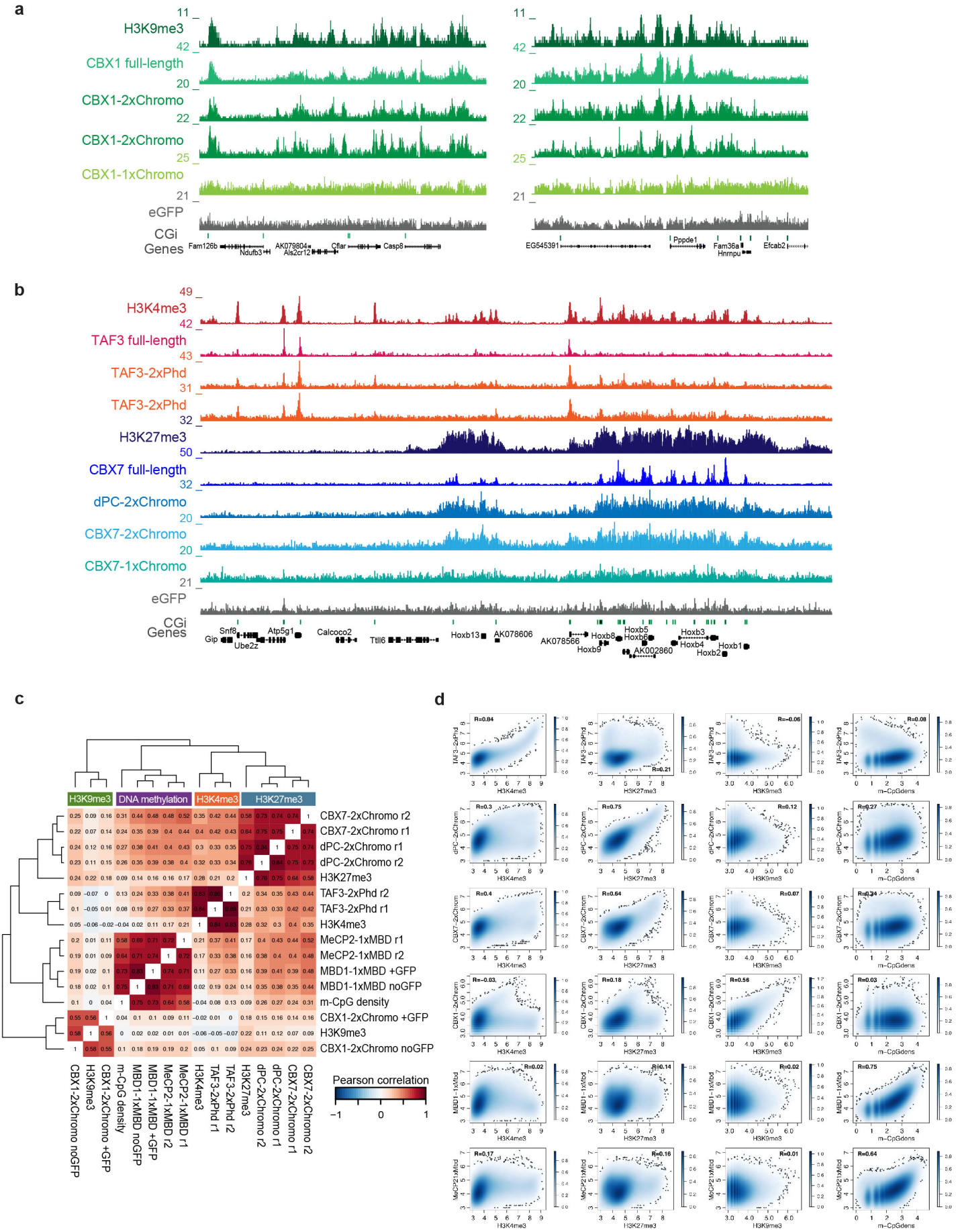
Villasenor et al. Supplementary Figure 1



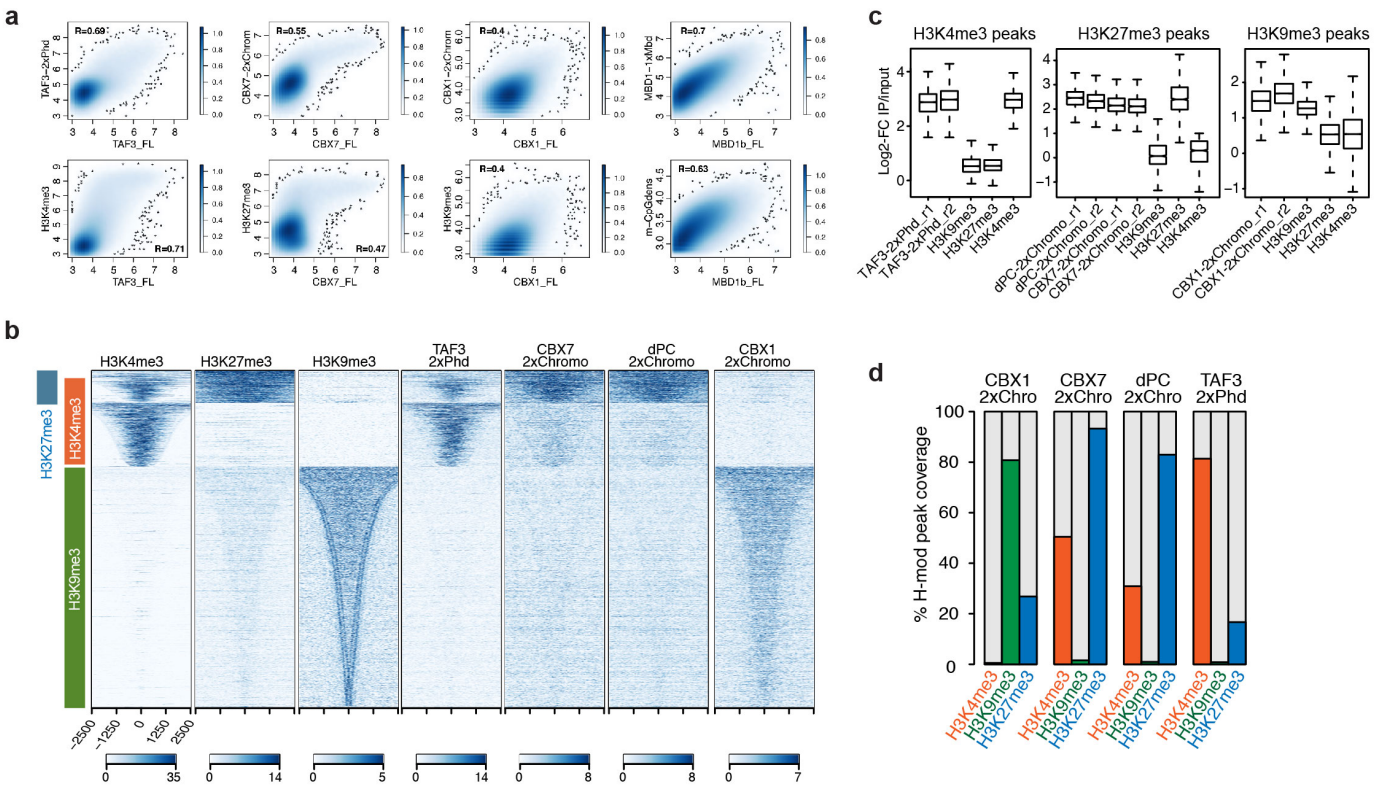
Villasenor et al. Supplementary Figure 2



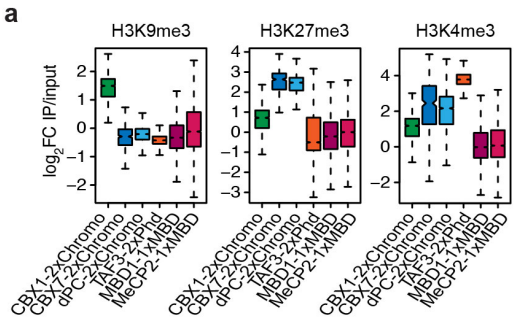
Villasenor et al. Supplementary Figure 3

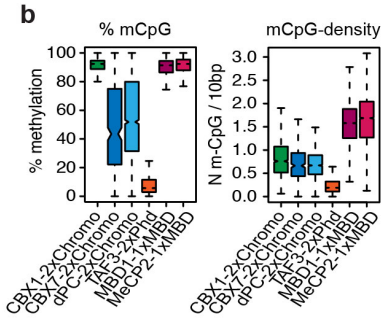


Villasenor et al. Supplementary Figure 4



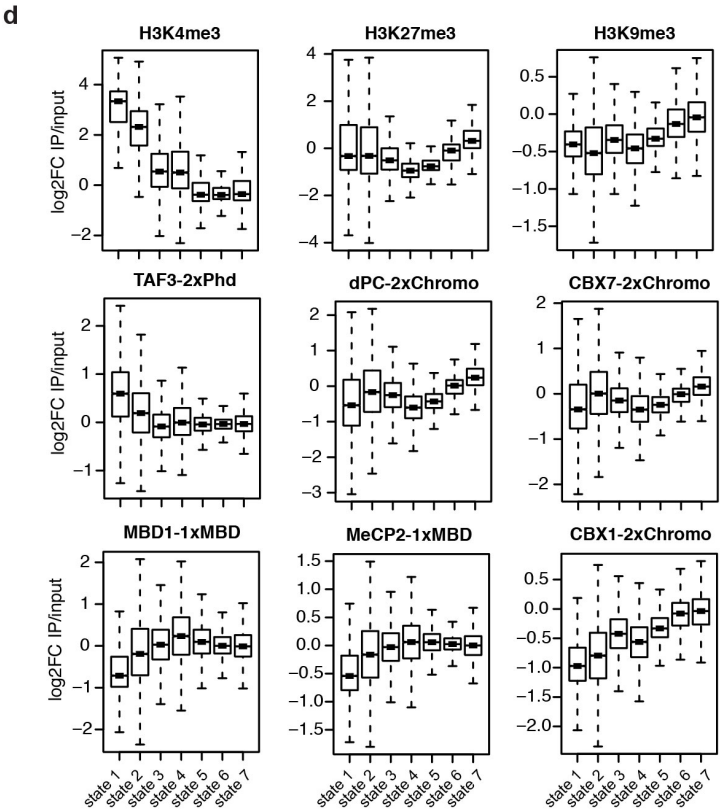
Villasenor et al. Supplementary Figure 5

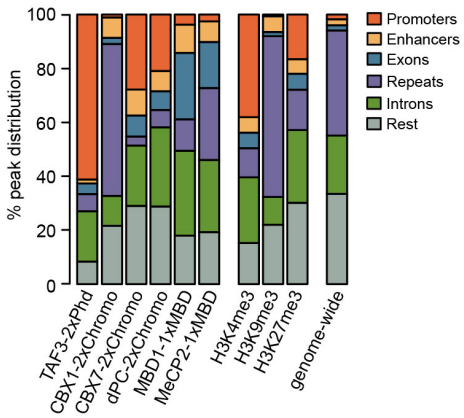


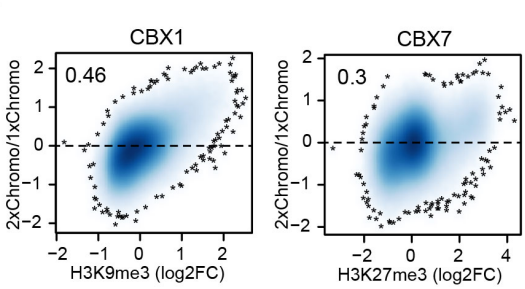


c

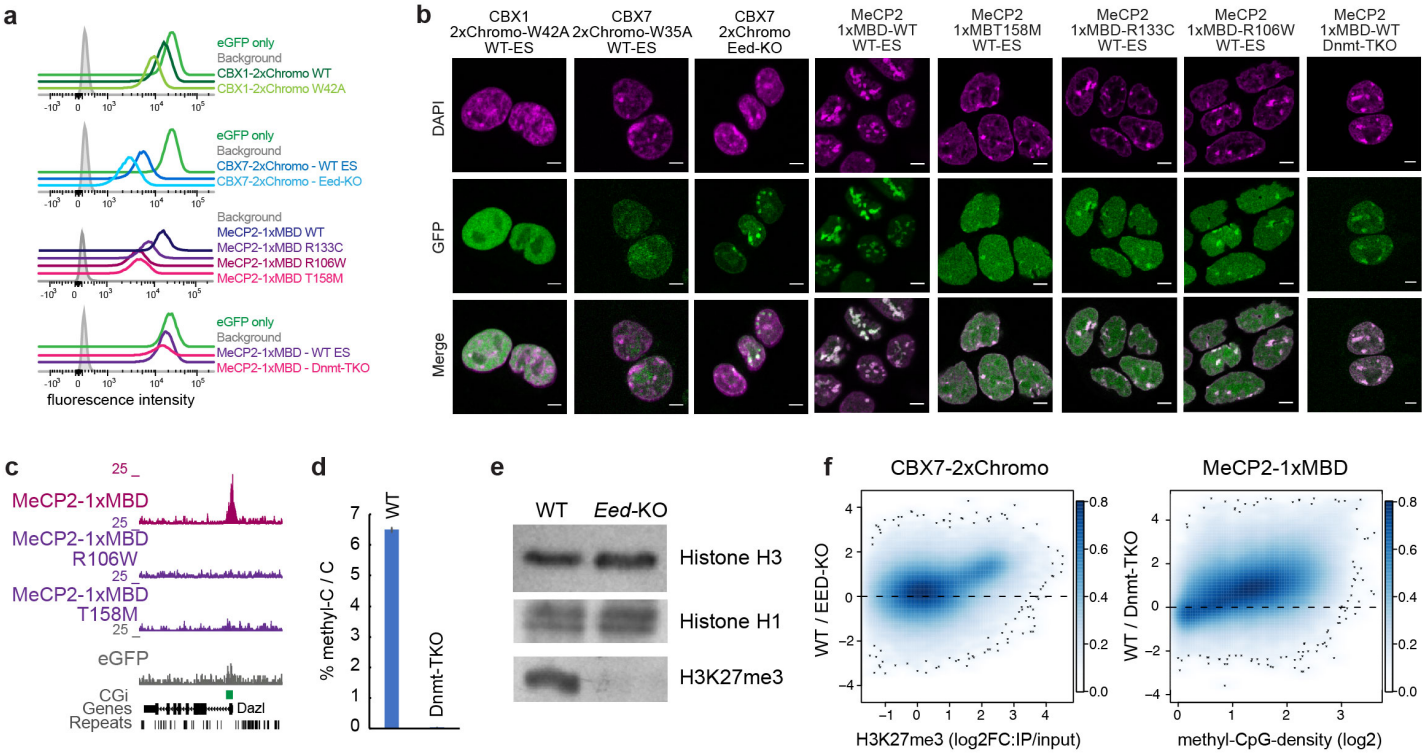
State	Feature	H3K27m3	H3K4m3	H3K4m1	H3K36m3
1	K4m3	0.07	0.92	0.05	0.03
2	K4m1/3	0.17	0.85	0.88	0.05
3	K4m1	0.01	0.01	0.47	0.02
4	K4m1+K36m3	0.01	0.05	0.59	0.71
5	K36m3	0.00	0.00	0.01	0.42
6	Unmarked	0.01	0.00	0.00	0.00
7	K27m3	0.29	0.00	0.02	0.00



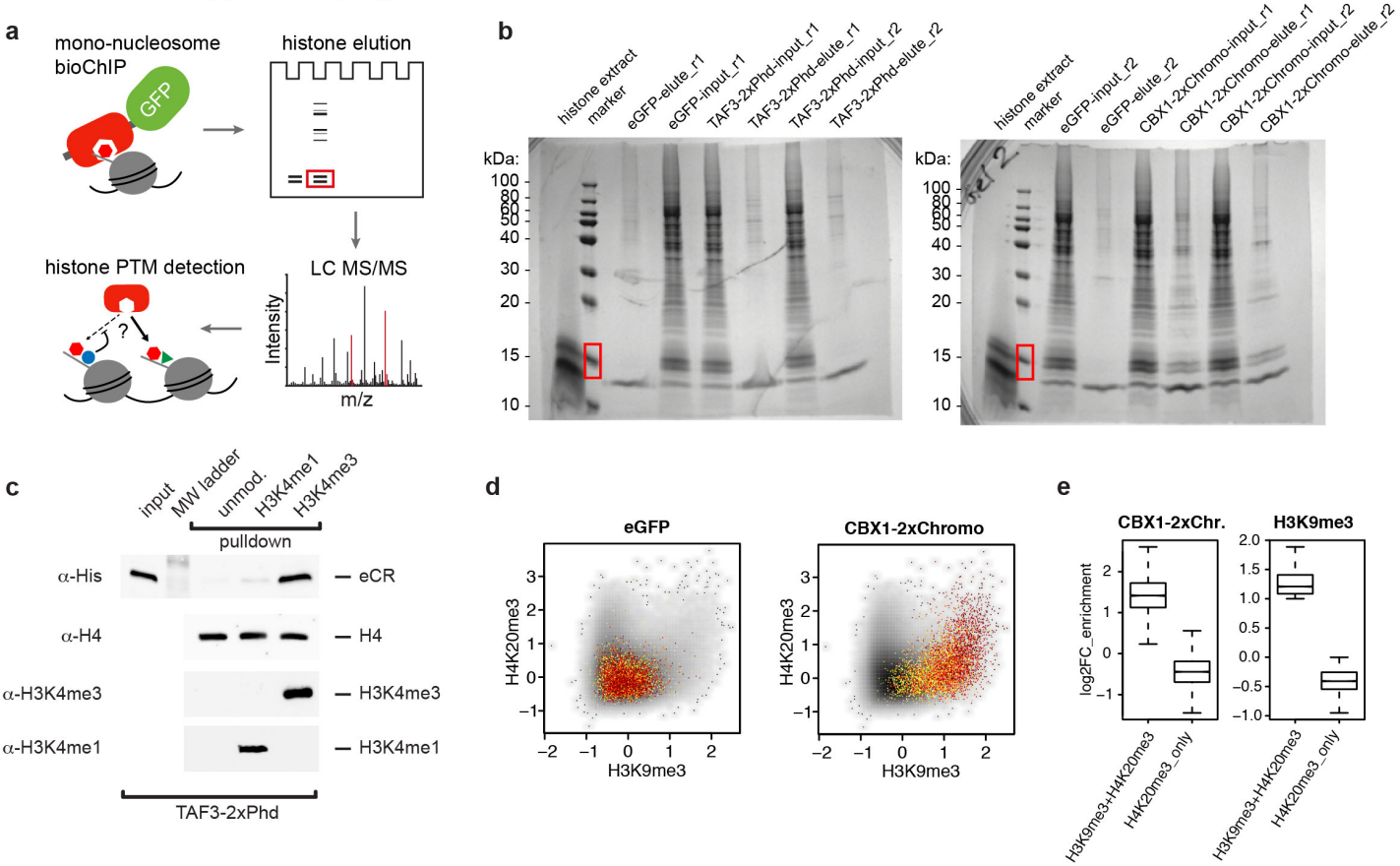




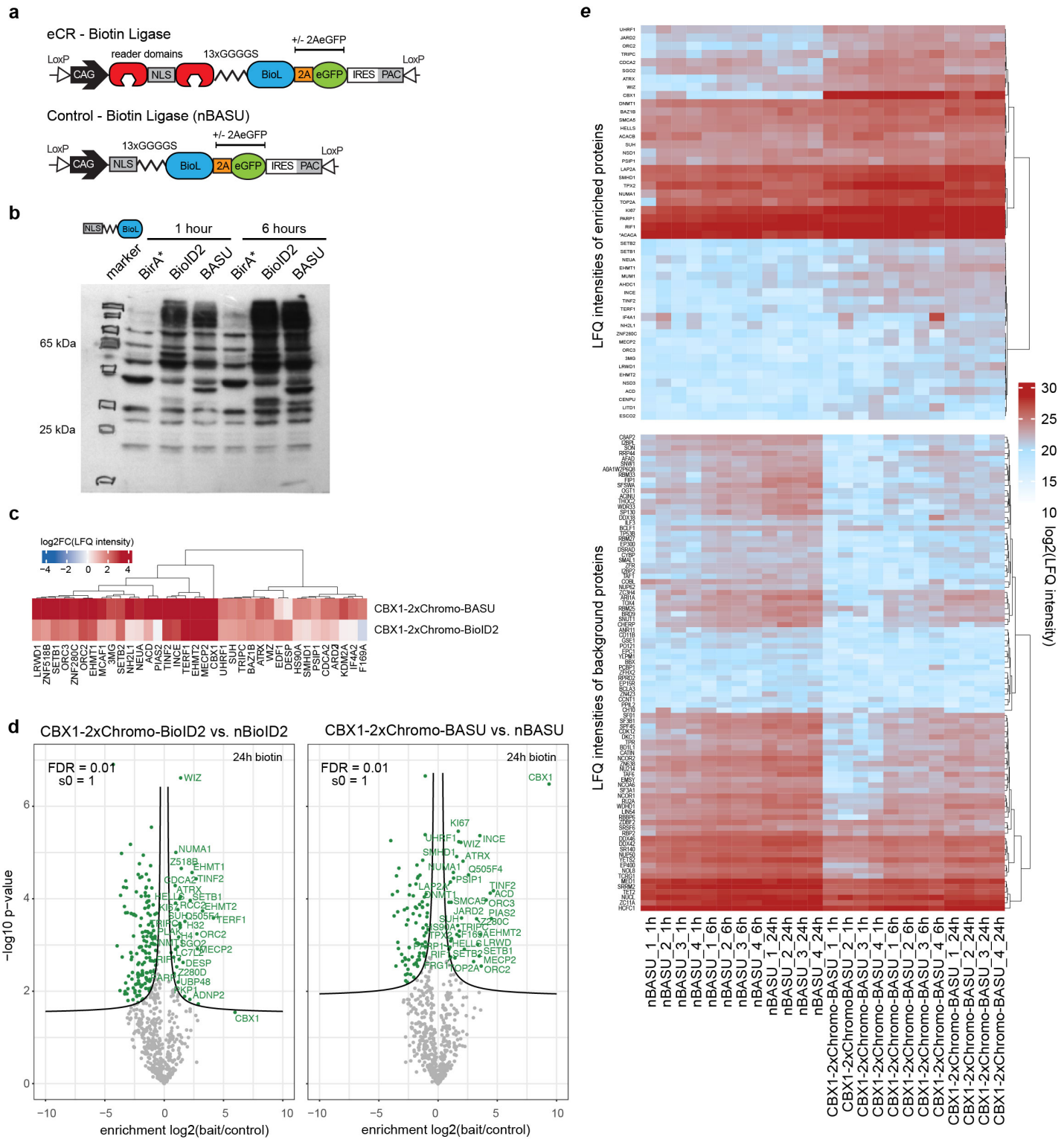
Villasenor et al. Supplementary Figure 6



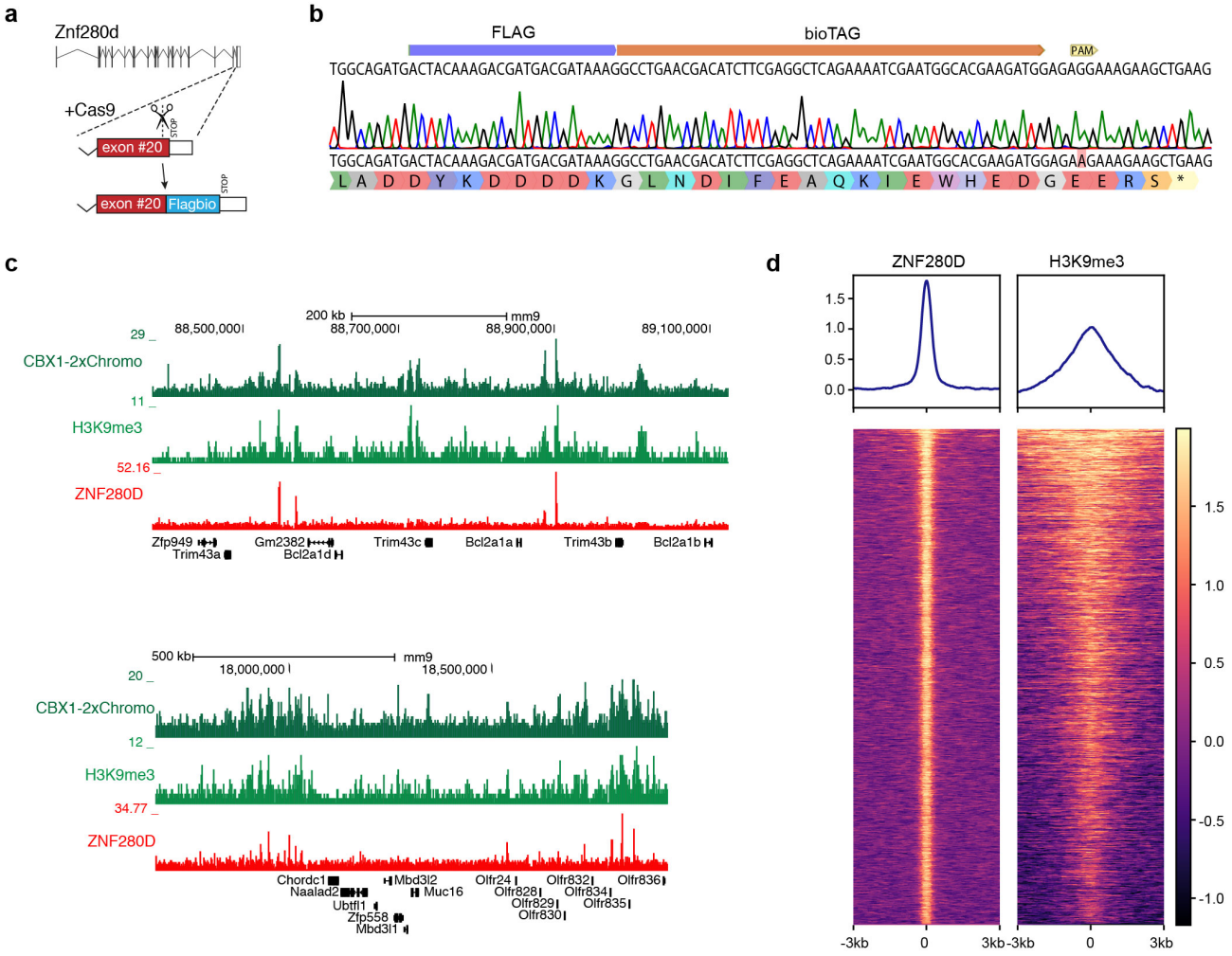
Villasenor et al. Supplementary Figure 7



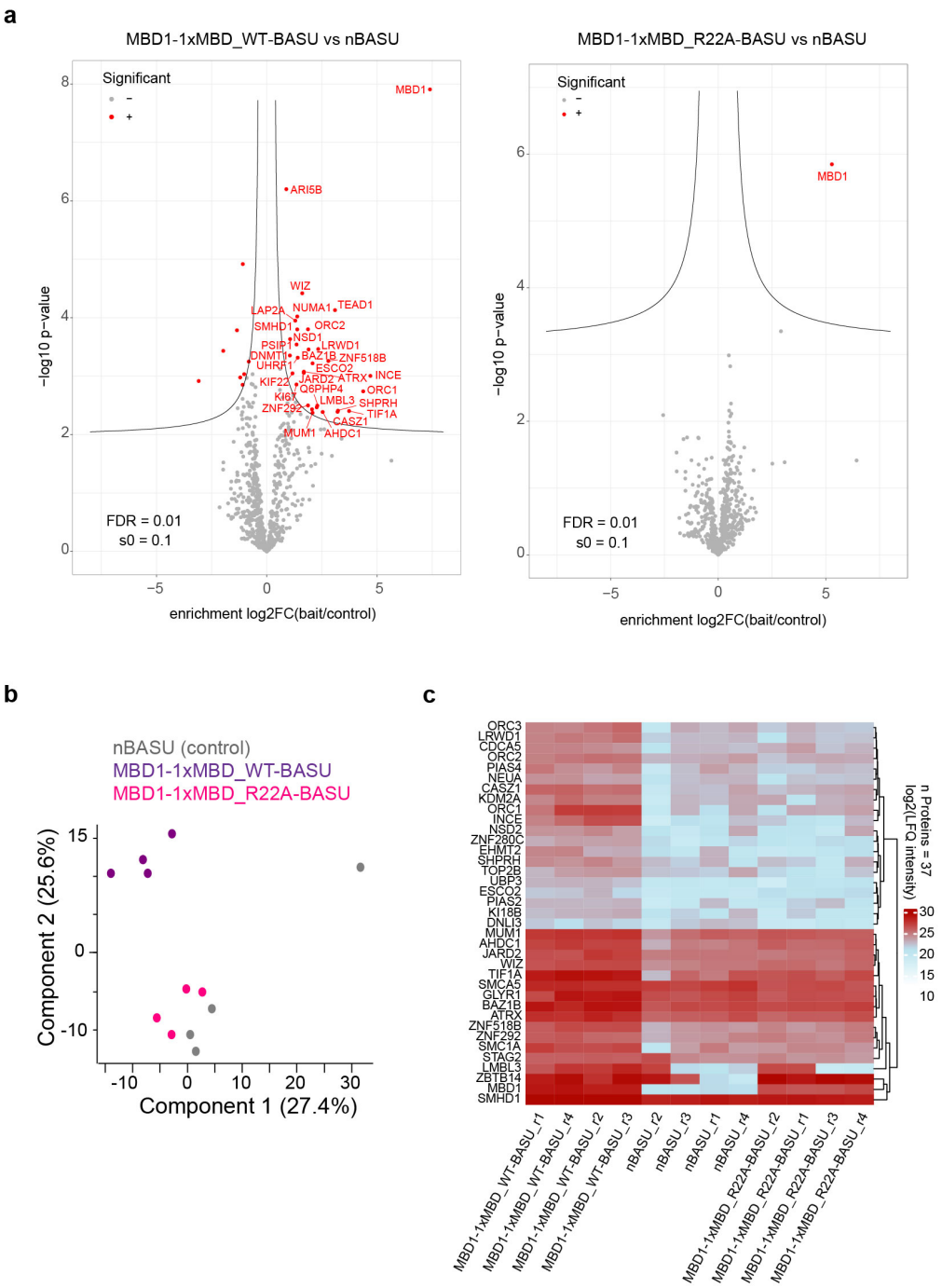
Villasenor et al. Supplementary Figure 8



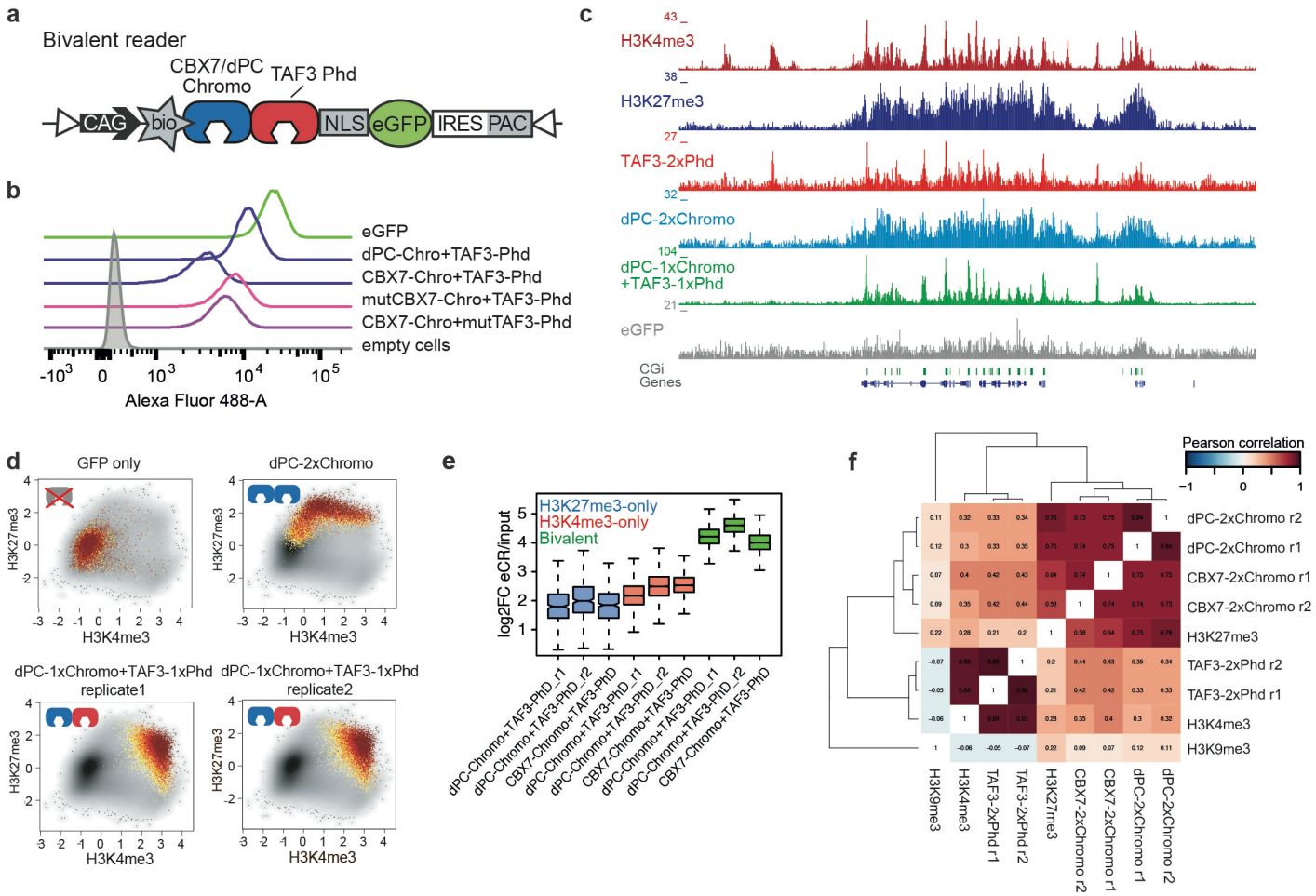
Villasenor et al. Supplementary Figure 9



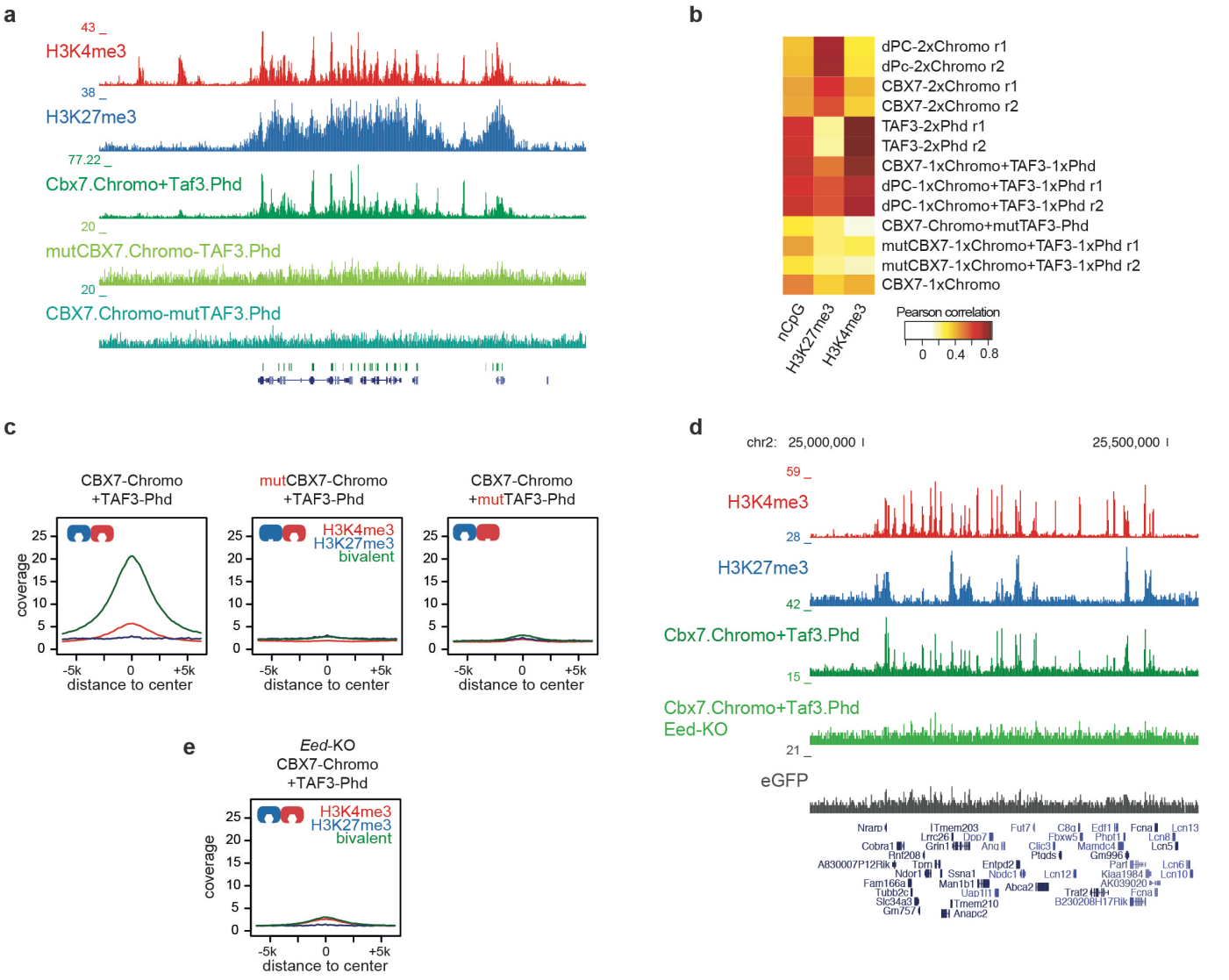
Villasenor et al. Supplementary Figure 10



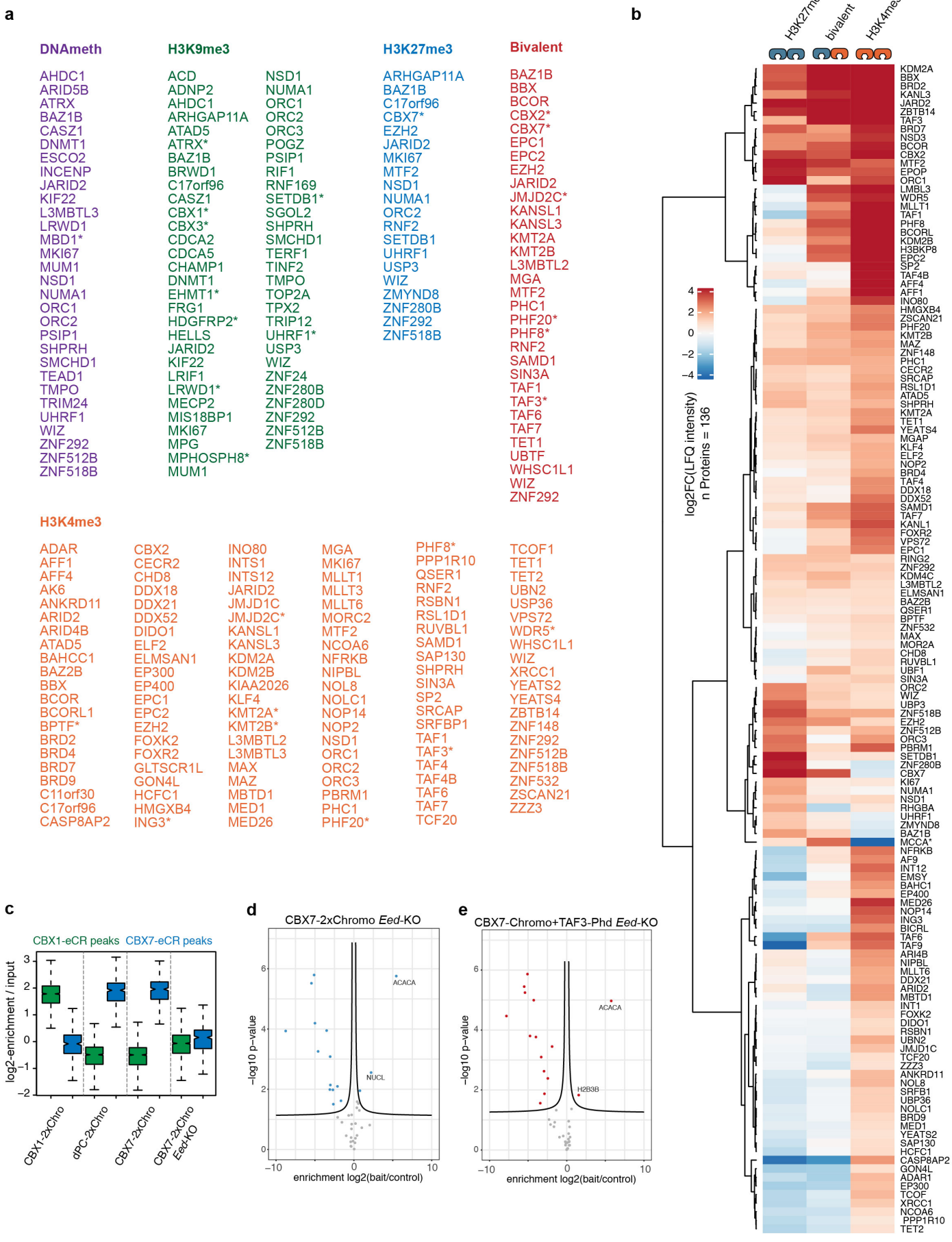
Villasenor et al. Supplementary Figure 11



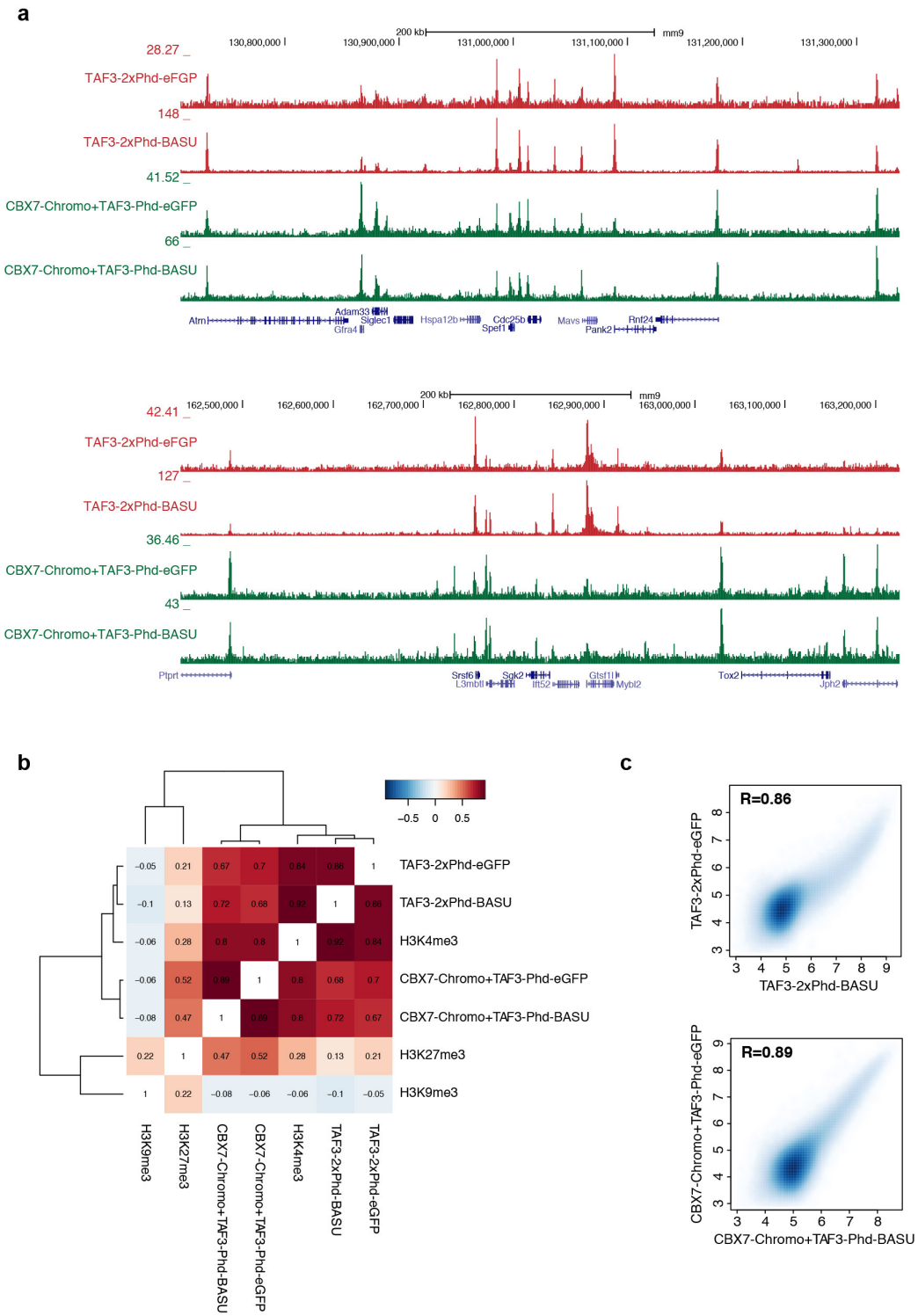
Villasenor et al. Supplementary Figure 12



Villasenor et al. Supplementary Figure 13



Villasenor et al. Supplementary Figure 14



Villasenor et al. Supplementary Figure 15